# Design and Optimization of Imaging Systems by Engineering the Pupil Function

by

Saeed Bagheri

Submitted to the Department of Mechanical Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Mechanical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2007
[June 2007]



Author .....
Department of Mechanical Engineering
May 3, 2007

Certified by .....
Daniela Pucci de Farias
Assistant Professor, Mechanical Engineering
Thesis Supervisor

Accepted by .....
Lallit Anand
Chairman, Department Committee on Graduate Students

# Design and Optimization of Imaging Systems by Engineering the Pupil Function

by

Saeed Bagheri

Submitted to the Department of Mechanical Engineering
on May 3, 2007, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Mechanical Engineering

## Abstract

It is expected that the ability to accurately and efficiently design an imaging system for a specific application will be of increasing importance in the coming decades. Applications of imaging systems range from simple photography to advanced lithography machines. Perhaps the most important way to make an imaging system meet a particular purpose is to engineer the pupil function of the imaging system. This includes designing a pupil surface and often involves the simultaneous design of a post-processing algorithm. Currently these design processes are performed mostly by using numerical optimization methods. Numerical methods in general have many drawbacks including long processing time and no guarantee that one has reached the global optimum. We have developed analytical approaches in designing imaging systems by engineering the pupil function.

Two of the most important merit functions that are used for the analysis of imaging systems are the modulation transfer function (MTF) and the point spread function (PSF). These two functions are standard measures for evaluating the performance of an imaging system. Usually during the design process one finds the PSF or MTF for all the possible degrees of freedom and chooses the combination of parameters which best satisfies his/her goals in terms of PSF and MTF. In practice, however, evaluating these functions is computationally expensive; this makes the design and optimization problem hard. In particular, it is often impossible to guarantee that one has reached the global optimum.

In this PhD thesis, we have developed approximate analytical expressions for MTF and PSF of an imaging system. We have derived rigorous bounds on the accuracy of these expressions and established their fast convergence. We have also shown that these approximations not only reduce the calculation burden by several orders of magnitude, but also make the analytic optimization of imaging systems possible. We have studied the detailed properties of our approximations. For instance we have shown that the PSF approximation has a complexity which is independent of certain system parameters such as defocus. Our results also help in better understanding the behavior of imaging systems. In particular, using our results we have answered a fundamental question regarding the limit of extension of the depth of field in imaging systems by pupil function engineering. We have derived a theoretic bound and we have established that this bound does not change with change of phase of pupil function. We have also introduced the concept of conservation of spectral signal-to-noise ratio and discussed its implications in imaging systems.

Thesis Supervisor: Daniela Pucci de Farias
Title: Assistant Professor, Mechanical Engineering

# Acknowledgments

First, I would like to thank my adviser, Prof. Daniela Pucci de Farias. This work would not have been possible without her enthusiasm, guidance and generosity in allowing me to pursue my own ideas and research directions. I will forever be grateful to her for giving me increasing levels of responsibility as my graduate experience progressed. I would also like to thank my informal adviser, Prof. George Barbastathis. I am deeply indebted to him for being an enthusiastic and genius supervisor. His comments and insights were invaluable sources of support for me. I am also grateful to the other member of my thesis committee, Prof. Peter So for his insightful recommendations.

The research presented in this dissertation is a result of close collaboration with Prof. Mark A. Neifeld and Drs. Paulo E. X. Silveira, Ramkumar Narayanswamy and Edward Dowski. I would like to thank them all for their valuable comments and suggestions. Special thank goes to Paulo for his continuous support and encouragement and many hours of helpful discussion. I am also grateful to researchers in CDM Optics for two great summers in the research industry. Furthermore, I would like to thank Drs. Michael D. Stenner, Kehan Tian and Wenyang Sun for insightful discussions and exchange of ideas.

My education at MIT was enriched by interaction with fellow students and friends. I am especially grateful to Angelina Aessopos, Mohammad-Reza Alam, Husain A. Al-Mohssen, Mohsen Bahramgiri, Lowell L. Baker, Reza Karimi, Hemanth Prakash, Judy Su, Dimitrios Tzeranis and Yunqing Ye for making the courses and research as well as life in MIT much more fruitful. I would like to also thank my other friends, namely Salman Abolfathe Beikidezfuli, Mohammad Araghchini, Mohammad Azmoon, Hoda Bidkhori, Megan Cooper, Eaman Eftekhary, Shaya Famini, Ali Farahanchi, Ghassan Fayad, Julia Greiner, MohammadTaghi Hajiaghayi, Fardad Hashemi, Ali Hosseini, Asadollah Kalantarian, Ali Khakifirooz, Danial Lashkari, Hamed Mamani, Vahab Mirrokni, Ali Parandehgheibi, Mohsen Razavi, Saeed Saremi, Ali Tabaei, Hadi Tavakoli Nia, Luke Waggoner and many others for creating a joyful and unforgettable three years for me. I am also grateful to the staff of Mechanical Engineering Department, specially Leslie Regan, Joan Kravit and Laura E. Kampas for providing a nice academic environment.

I would like to use this opportunity to thank my undergraduate advisers, Prof. Ali Amirfazli, Prof. Mikio Horie and Dr. Daiki Kamiya for shaping my career path. Also I am grateful to Seid Hossein Sadat, Hamidreza Alemohammad and Reza Hajiaghaee Khiabani for many helpful discussions.

I thank my family for their unconditional support throughout these years. They provided everything I needed to succeed. My parents' everlasting love and prayer has always been and will remain a primary source of support for me and I am forever indebted to them for that. I am grateful to my brother, Vahid for being out there for me and to my sister, Maryam for her love. Last but not least, I thank my wife, Mariam for her understanding, inspiration and endless love. This thesis is dedicated to my family.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction and Overview

One of the most important and perhaps first applications of the field of optics is imaging. Imaging systems have developed and got more and more complicated throughout the history. Each of these added complications has enhanced the performance of imaging systems in some way. These range from inventing new devices for recording the image to increasing the resolution of imaging systems to image ultra-small structures like atomic-level roughness of surfaces. In this thesis we focus on one of such added complications: pupil function engineering.

## 1.1 Pupil Function Engineering

Pupil function engineering involves modification of the wavefront so that the imaging system meets a particular purpose. This modification often happens in the pupil plane, hence the name pupil function engineering. The pupil function can be considered as a complex function which is to be multiplied by the original wavefront at the pupil plane. This function has maximum amplitude of unity; in this thesis we only consider pupil functions which preserve the light collected by the imaging system; i.e. pupil functions that have amplitude of exactly equal to unity. This ensures that there is no absorption in the pupil plane.

Based on the above discussion, pupil function engineering in our context, consists of designing the phase function only. Note that pupil function is a function of pupil plane coordinates. So,

strictly speaking, one can think of our final designed pupil function as a physical object. This object (compare with contact lens that you wear everyday) modifies the phase of the incoming light and preserves the amplitude. It is clear that choices for pupil function is endless and that is precisely what makes the problem of pupil function engineering interesting. It should be added that pupil function engineering is often accompanied with a proper post-processing algorithm to get the final image. Although we do not cover the details of designing post-processing algorithms we do take into consideration the fact that post-processing algorithms exist. In fact, the final designed pupil function has to satisfy specific requirements to allow the use of post processing algorithms.

The ability to accurately and efficiently design an imaging system for a specific purpose is of increasing importance in the coming decades. Applications of imaging systems that have used pupil function engineering for this goal range from simple photography to advanced lithography machines. To mention some of the current application fields for pupil function engineering, we can name adaptive optics, phase retrieval, aberration correction, photography, general microscopy, optical lithography, integral imaging and computational tomography among others.

In this thesis, we study pupil function engineering, its tools, applications and limits from an analytic point of view. In Section 1.2 we motivate our approach by reviewing some the current problems in pupil function engineering and some of the potential benefits of our approach. In Section 1.3 we present the recent results in the literature related to this thesis and re-introduce our work in this context. Finally, in Section 1.4 we briefly mention our contributions in this thesis.

## 1.2 Motivation

The process of pupil function engineering, like any other design method has two main phases: (i) stating the imaging system requirements in the proper language, and (ii) finding the design parameters so that the imaging system meets those requirements. The challenge, however, seats in the transition from first phase to the second phase.

### 1.2.1 Computation Time Versus Accuracy

In imaging system design, there are a few different ways to mathematically state the imaging system requirements. Usually this is done by translating the system requirements to the language of either impulse response or frequency response of imaging systems. Two of the most important functions that are used for the analysis of imaging systems are the modulation transfer function (MTF) as a measure of frequency response and the point spread function (PSF) as a measure of the impulse response.

The process of transition from the MTF or the PSF to the final designed pupil function involves many instances of calculating PSF and/or MTF. In particular, depending on the number of possible degrees of freedom in final imaging system, the number of MTF and/or PSF calculation grows exponentially. As it is discussed in Chapters 2 and 3, based on a given pupil function, the calculation of MTF or PSF, each involves solving a double integral. This hints the process of pupil function engineering is very computationally expensive.

As it was mentioned in the previous Section, in pupil function engineering we are trying to design a phase value as a function of pupil coordinate and thus the complexity of final designed imaging system can be arbitrarily large. This increases the potential computational burden of the pupil function engineering even further.

This computational expense motivates us to look for a *better* way to relate the MTF and the PSF to the pupil function. In particular we are interested to find a relationship that allows us to move from pupil function to MTF and PSF fast and accurately. We view this as a major tool in pupil function engineering. In this thesis we have developed an approximate analytical relation between the MTF and the pupil function and the PSF and the pupil function.

### 1.2.2 Analytical Optimizations

In each iteration of pupil function engineering, the MTF or the PSF are calculated based on a given pupil function. This pupil function is, in turn, a result of a set of chosen design parameters. During the process of pupil function engineering the above iteration is repeated for many different design

parameters and in the end, a set of designed parameters is chosen as the optimal set. This set of design parameters is optimal in the sense that the MTF or the PSF resulted from the corresponding pupil function matches the stated MTF or PSF requirements the best.

An instant question that can be raises is what do we know about optimality of the final pupil function. For instance we would like to know if our result is the global optimum or not. In fact, in finite iterations of the above algorithm we cannot guarantee anything about global optimality.

An alternative to the above approach is to use analytic optimization algorithms; however to do so we need to have a closed form expression relating the MTF and the PSF to the pupil function. It is well-known that such a relation does not exist except in the integral form and thus this motivates us to develop analytic approximate expressions relating the MTF and the PSF to the pupil function.

### 1.2.3 Physical Understanding

Due to the high complexity of pupil function engineering, usually the optimization process and result carry no or little intuition about the physics of the problem. In particular, there are many classes of problems that pupil function engineering is known to be well-suited for; nevertheless the physical limits of pupil function engineering for most of these problems has not been studied and is not known.

For instance, consider the problem of extension of depth of field. It is known that pupil function engineering can extend the depth of field. There has been many instances of successful imaging system designs that have used this. Yet, there are some unanswered fundamental questions in this regard. One of the fundamental questions in this regard is: to what extent one can extend the depth of field of an imaging system using design and optimization of the pupil function? Questions of this nature, motivate us to study the pupil function engineering analytically. Also, during the optimization process itself, having an analytic expression for the problem can help the designer a lot as to what are the important parameters, what are the sensitive parameters, etc.

## 1.3 Background

In this section we perform a quick review of results related to pupil function engineering. In particular, we present results related to both numerical and analytical approaches in pupil function engineering. Finally, we consider the depth of filed and results related to extension of depth of field.

### 1.3.1 Numerical Results

Traditionally calculation of MTF and PSF using a specific pupil function has been done numerically. One way of doing this is to directly calculate the finite Riemann sum as an approximation to the final integral. It is a common practice to use fast Fourier transform (FFT) rather than direct integration to enhance the speed of calculation. However, the FFT method also fails to perform well as the resolution of interest or the accuracy of interest increases. In what follows we review the performance of FFT for calculation of PSF using a given pupil function. Note that MTF and PSF are related using Fourier optics and thus the same discussion applies to MTF as well.

Here we quickly review the trade off between accuracy and computational expense in the FFT method [1, 2]. In imaging systems we are usually interested in two-dimensional FFTs. The number of calculations necessary for a two-dimensional FFT of an $\mathcal{N} \times \mathcal{N}$ array is $2\mathcal{N}^2 \log \mathcal{N}^2$. Thus the time necessary is proportional to this expression too. In computational imaging systems we take the FFT of the pupil function to get PSF. Accuracy of FFT highly depends on the complexity of the pupil function. This is because as this complexity increases the fixed number of sample points fail to capture the complexity of pupil function well.

This has serious drawbacks in optimization algorithms. First, during each iteration the complexity of pupil function changes and thus to keep accuracy fixed we need to change the FFT size appropriatly. This makes the optimization algorithm more complicated.

Secondly, as it was shown above the number of sample point and thus the calculation burden grows very fast with the complexity of pupil function. In particular, this limits our design abilities to incorporate more complicate pupil functions in imaging systems.

### 1.3.2 Analytical Results

There has been many efforts to analytically approximate important functions in imaging systems. For the same reason as last Section, we only consider PSF and methods to analytically approximate this function. The original Nijboer-Zernike method that can be applied to very simple pupil function has very limited application. [3] Recently, extensions of the original Nijboer-Zernike method have been developed in order to make it applicable to more complicated pupil functions. These expansions lead to a representation of the PSF whose complexity increases at least linearly with defocus [4, 5, 6, 7]. This means given a class of pupil functions, as the value of some of the parameters increases, the complexity of the calculation increases too.

### 1.3.3 Depth of Field

A common problem encountered in the design of imaging systems consists of finding the right balance between the light gathering ability and the depth of field (DOF) of the system. Having high signal-to-noise ratio (SNR) at the detector of an imaging system over a large range of depths of field has been the utmost goal in many imaging system designs [8, 9, 10].

Traditionally, most of the attention in the literature has been focused on increasing the depth of field for special problems of interest. This typically includes cases of successfully designed imaging systems that work in an extended depth of field. Usually in these systems SNR is shown to be within acceptable limits depending on the particular goal. There are however cases in which a subclass of design problems (for instance, cubic-phase pupil function) are studied analytically where the limits of extension of depth of field in terms of generic acceptable SNR is discussed more rigorously [11, 12, 9, 13, 14].

Traditionally (as we have all experienced with our professional cameras) one can extend the depth of field by controlling the aperture stop size. Albeit very simple, this method has serious drawbacks such as significantly reducing the optical power available and the highest spatial frequency [15]. This limits the practical use of this method to very short ranges of depth of field [16]. Pupil function engineering combines aspheric optical elements and digital signal processing to extend the depth

of field of imaging systems. [17, 18, 19, 20]. Although numerous important industrial problems are solved using pupil function engineering, there is no concrete statement about the extent pupil function engineering can improve SNR over the depth of field of interest.

## 1.4 Contribution

In this thesis, we study pupil function engineering, its tools, applications and limits from an analytic point of view. In particular, we derive approximate analytic expressions for the MTF and the PSF. Using our expressions one can save a lot of computational power at a practically negligible accuracy expense. We solve some optimization problems using our expressions. We also answer the fundamental question regarding the extension of depth of field in imaging system.

In Chapter 2, we derive an approximate analytical expression for the MTF of an imaging system possessing a shifted cubic phase pupil function. We derive the error bounds of our approximation and establish its high accuracy (see Theorem 2.2.1). Using the approximate representation of the MTF we solve the problem of extension of depth of field analytically for two cases of interest: *uniform quality* imaging and *task-based* imaging. We also show how the analytical solutions given in this Chapter can be used as a convenient design tool as opposed to previous lengthy numerical optimizations.

In Chapter 3, we introduce a new method for analyzing the diffraction integral for evaluating the PSF. Our approach is applicable when we are considering a finite, arbitrary number of aberrations and arbitrarily large defocus simultaneously. We present an upper bound for the complexity and the convergence rate of this method (see Theorem 3.5.1). We also compare the cost and accuracy of this method to traditional ones and show the efficiency of our method through these comparisons. This has applications in several fields that use pupil function engineering such as biological microscopy, lithography and multi-domain optimization in optical systems.

In Chapter 4, we discuss the limit of depth of field extension for an imaging system using pupil function engineering. In particular we consider a general imaging system in the sense that it has arbitrary pupil-function phase and we present the trade-off between the depth of field of the system

and the spectral SNR over an extended depth of field. Using this, we rigorously derive the expression for the tightest upper bound for the minimum spectral SNR, i.e. the limit of spectral SNR improvement (see Theorem 4.4.1). We also draw the relation between our result and the conservation of brightness theorem and establish that our result is the spectral version of the brightness conservation theorem. Finally, we conclude in Chapter 5.

# Chapter 2

# Modulation Transfer Function

In this Chapter we derive an approximate analytical expression for the modulation transfer function (MTF) of an imaging system possessing a shifted cubic phase pupil function. This expression is based on an approximation using Arctan function. Using the approximate representation of the MTF we solve the problem of extension of depth of field analytically for two cases of interest: *uniform quality* imaging and *task-based* imaging. We derive the optimal result in each case as a function of the problem specification. We also compare the two different imaging cases and discuss the advantages of using our different optimization approach for each case. We also show how the analytical solutions given in this Chapter can be used as a convenient design tool as opposed to previous lengthy numerical optimizations.

## 2.1 Introduction

Pupil function engineering is a computational imaging technique used to greatly increase imaging performance while reducing the size, weight, and cost of imaging systems [18]. It consists of the combined use of optical elements with digital signal processing in order to better perform a required imaging task. For example, pupil function engineering can be used to extend the depth of field of an imaging system [17, 21]. In traditional (without pupil function engineering) imaging systems, such an extension of the depth of field is usually obtained by reducing the aperture stop, consequently

reducing the resolution and light gathering capacity of the imaging system. Because pupil function engineering elements typically are non-absorbing phase elements, they allow the exposure and illumination to be maintained while producing the depth depth of field of a slower system [18, 20].

A challenging process in designing systems with extended depth of field is choosing the right phase element. The design goal is to make the point spread function (PSF) of the optical system defocus-invariant; i.e. to make the PSF of the optical system shape-invariant as the object moves along the desired depth of field. Having a defocus-invariant PSF facilitates the image reconstruction using a simple deconvolution filter. Simultaneously, one tries to keep the MTF as high as possible as the object moves along the desired depth of field. This is done, in order to transfer the most information possible from the object to the optical sensor. Here, by the most information possible we refer to the space bandwidth product of the imaging system or in other words the maximum number of resolvable spots [22, 23, 24].

One of the phase elements that is most often used in practice is described by a cubic phase, $\Phi(\hat{x}, \hat{y})$, expressed as

$$\Phi(\hat{x}, \hat{y}) = \alpha \left[ (\hat{x} + \delta)^3 + (\hat{y} + \delta)^3 \right].$$

where $\alpha$ is the cubic phase coefficient, $\delta$ is the cubic phase shift and $(\hat{x}, \hat{y})$ are the normalized Cartesian coordinates at the pupil plane.

This phase surface has some interesting properties; among which is the fact that the PSF of the optical system which is equipped with this phase element is defocus-invariant. This property makes this phase element an excellent choice for the problem of extension of the depth of field. Having chosen this type of phase element the design usually consists of numerically maximizing the MTF of the optical system. The optimization parameters are phase element parameters (e.g. $\alpha$) and optical system parameters (e.g. distance between the image plane and the exit pupil) [24].

This process of numerical design and optimization is lengthy and time consuming, for one needs to numerically evaluate the MTF for every variation of design parameter values, and a large number

of parameter values have to be visited. Furthermore this process needs to be redone for every and each specific problem. On top of all this, there is no theoretical guaranty that one is actually reaching the global optimum design within a limited time of numerical optimization [25, 26].

In this Chapter we offer an alternative to numerical optimization by modeling and solving the design problem analytically. This is mainly a result of our developed approximation to MTF. We derive the expression for a generalized MTF with cubic phase element in pupil plane. In this model we assume a diffraction-limited lens, an incoherent illumination and a cubic-phase element. We perform an accuracy analysis and show that the developed approximation has a very good accuracy (97% on average) in the regions of interest in imaging design.

This model provides us with the MTF as a function of defocus. This generalized MTF is then used to optimize the imaging system. We analytically solve the cubic phase element design and optimization problem for two general imaging problems. These two problems are: (i) to extend the depth of field for uniform image quality imaging systems (e.g. normal photography, cellphone cameras, etc. ) and (ii) to extend the depth of field for task-based imaging systems (e.g. bar code reader, iris recognition, etc.).

In the next Section we derive the analytic MTF representation, which will be used as a basis for our optimization in the rest of the Chapter. In Section 2.3 we solve the problem of extending the depth of field for the case of imaging with uniform image quality. We go over the optimality criteria and we solve the optimization problem analytically. We present an example and show how our results apply in solving a practical problem. In Section 2.4, we solve the problem of extending the depth of field for task-based imaging. In Section 2.5 we compare both results from Sections 2.3 and 2.4 a nd discuss their relative benefits. We also go over some of the general results that could be deduced from the optimal solution graphs and their applications to design problems.

## 2.2 Analytic MTF

In this Section we derive an analytic approximation for the MTF of an imaging system with a cubic phase element installed in its aperture. We assume an imaging system with circular aperture being

illuminated with incoherent light. Figure 2-1 shows a schematic view of our optical system. Using simple Fourier optics one can get the expression for MTF of such optical system. Note that the ultimate goal of this chapter is to maximize MTF. However there is a fundamental limit to that due to the conservation of ambiguity function. It has been observed that generally the most efficient way of managing this limit is to maximize MTF only on two orthogonal axis, thus keeping the used portion of this fixed area as small as possible [27, 28, 29]. Please see Appendix A or Chapter 4 for more detailed discussion. Due to the symmetry of problem, it suffices to analyze MTF in any of these two orthogonal directions. Thus, we can continue with the revised version of the MTF equation as below

$$MTF^e(u,0) = \frac{1}{\pi}\left|\int_{-y_m}^{y_m}\int_{-x_m}^{x_m} e^{ki\left[(4W_{20}u)\,\hat{x}+(6\alpha u)\,\hat{x}^2\right]}\,d\hat{x}\,d\hat{y}\right|, \tag{2.1}$$

where $MTF^e$ is the exact value of MTF, $\hat{x}$ and $\hat{y}$ are normalized Cartesian coordinates of the pupil, $u$ and $v$ are normalized spatial frequencies in $\hat{x}$ and $\hat{y}$ directions respectively, $W_{20} = (D^2/8)\,(1/d_i + 1/d_o - 1/f)$ is the defocus coefficient and $\alpha$ is the cubic-phase coefficient; $k = 2\pi n/\lambda$, $f$, $d_i$, $d_o$ and $D$ are the wave number, imaging system focal length, distance from the image plane to the exit pupil, distance from the object plane to the entrance pupil and aperture diameter respectively. The last three parameter definitions are illustrated in Fig. 2-1. Note that $W_{20}$ has the dimension of length. Finally, $x_m$ and $y_m$ are defined as

$$\begin{aligned} x_m &= \sqrt{1-\hat{y}^2} - u, \\ y_m &= \sqrt{1-u^2}. \end{aligned} \tag{2.2}$$

The details of derivation of Eq. (2.2) is given in Appendix A. At this point we have two integrals which cannot be analytically evaluated in a closed form. In particular, for any set of imaging system parameters calculating the value of MTF requires numerical calculation of a double integral. This is

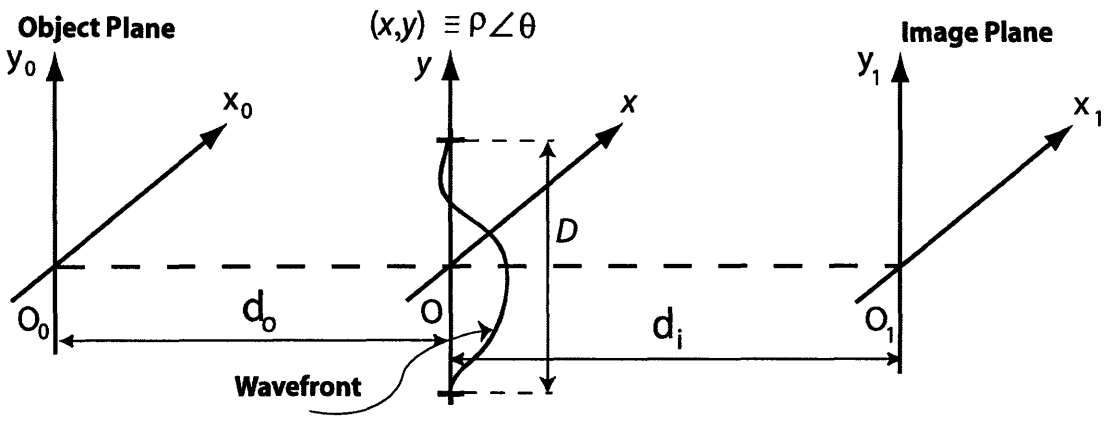



Figure 2-1: Schematic view of optical system under consideration.

often very expensive during a design process which requires many instances of MTF calculation. To overcome both of these problems (lack of closed form solution and computational cost), we introduce an approximation for MTF. Our results are based on the following novel approximation:

$$\int_0^X \exp(\mathrm{i}t^2)\mathrm{d}t \approx \frac{\mathrm{i}+1}{\sqrt{2\pi}} \operatorname{Arctan}\left(\sqrt{\pi}\,X\right), \tag{2.3}$$

Based on Eq. (2.3) we have deriven two different approximations for MTF. The first approximation, $MTF^{a1}$ is suitable for numerical calculation and the second approximation, $MTF^{a2}$, is suitable for analytic manipulations. The detailed derivation of these two approximations can be found in Appendices B and C, respectively. In the subsequent Sections we use $MTF^{a2}$ as our MTF approximation. The expression for $MTF^{a1}$ is too lengthy and thus is skipped. The expression for $MTF^{a2}$ is shown below

$$MTF^{a2}(u,0) = \frac{2}{3\pi^2 ku\alpha}\left[-\text{Arcsin}(u)+\sqrt{\frac{3\pi ku\alpha}{2}}\sqrt{1-u^2}\times \left(\text{Arctan}\left\{\sqrt{\frac{2\pi ku}{3\alpha}}\left[W_{20}+3\alpha(1-u)\right]\right\} -\text{Arctan}\left\{\sqrt{\frac{2\pi ku}{3\alpha}}\left[W_{20}+3\alpha(-1+u)\right]\right\}\right)\right]. \tag{2.4}$$

Equation (2.4) above is the approximate analytic expression for the MTF. Note that $u$ in this equation is the normalized spatial frequency (when $u = 1$, the system is at the diffraction limit). This expression for the MTF makes the analytic solution of an optical design problems that involves MTF mathematically tractable.

Before using any of these two approximations, we need to investigate their accuracy. The accuracy of $MTF^{a1}$ and $MTF^{a2}$ are studied in Appendices B and C, respectively. Here we present some of the results in this regard.

**Theorem 2.2.1.** *Let $\epsilon$ be the approximation accuracy and $\boldsymbol{C}$ be the sub-space of interest in design parameters space, such that*

$$\left|MTF^{e}(u,0) - MTF^{a1}(u,0)\right| \leq \epsilon(\hat{\alpha}, \hat{W}_{20}, u),$$

*and*

$$\boldsymbol{C} \equiv \{0.2 < u < 1\} \times \{2 < \hat{\alpha} < 10\} \times \left\{0 < \hat{W}_{20} < 8\right\}.$$

*Then, we have*

$$\max_{\boldsymbol{C}}\left\{\epsilon(\hat{\alpha}, \hat{W}_{20}, u)\right\} \leq 0.1,$$

$$\frac{1}{\|\boldsymbol{C}\|}\iiint_{\boldsymbol{C}} \epsilon(\hat{\alpha}, \hat{W}_{20}, u)\,\mathrm{d}\hat{\alpha}\,\mathrm{d}\hat{W}_{20}\,\mathrm{d}u \leq 0.03,$$

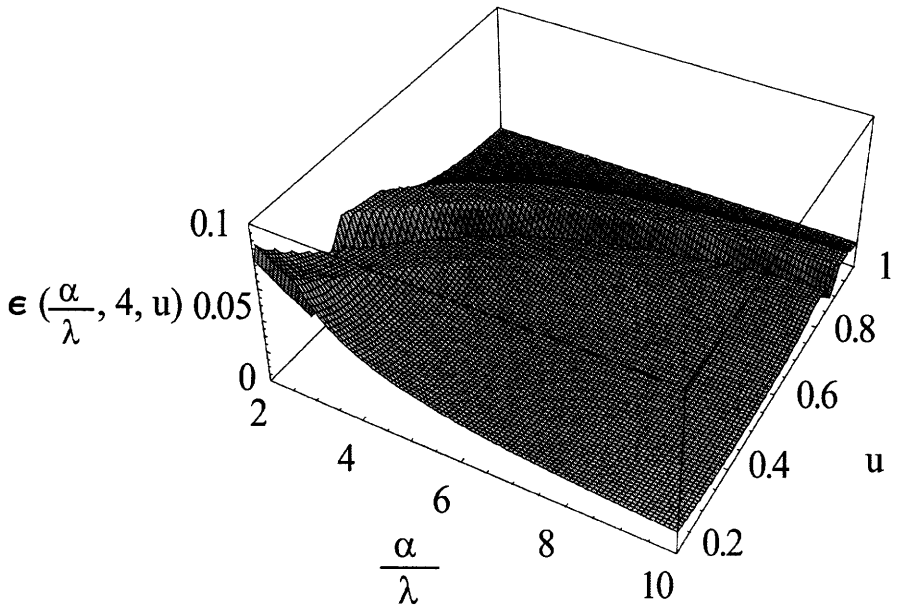



Figure 2-2: Plot of $\epsilon\left(\hat{\alpha}, 4, u\right)$ as a function of $\hat{\alpha}$ and $u$ (Note that all variables are dimensionless).

*and*

$$\left|MTF^{a2} - MTF^{a1}\right| \leq 0.1.$$

In Theorem 2.2.1, $\hat{\alpha} = \alpha/\lambda$ and $\hat{W}_{20} = W_{20}/\lambda$ are normalized cubic phase coefficient and normalized defocus coefficient, respectively. The function $\epsilon(\hat{\alpha}, \hat{W}_{20}, u)$ is a bound of our approximation error. The immediate interpretation of these results would be the high accuracy of $MTF^{a1}$. In particular, the minimum accuracy in $MTF^{a1}$ is 90% and the average accuracy is more than 97%. This establishes the practical usage of our approximation. Figure 2-2 shows the plot of $\epsilon(\hat{\alpha}, 4, u)$ as a function of $\hat{\alpha}$ and $u$. The other important result is regarding the accuracy of $MTF^{a2}$. As it is shown in Theorem 2.2.1 the difference between $MTF^{a2}$ and $MTF^{a1}$ is bounded. Figure 2-3 shows the plot of $\left|MTF^{a2} - MTF^{a1}\right|$ as $\hat{\alpha}$ and $u$ varies while $\hat{W}_{20} = 4$.

It should be noted that most of the results in Theorem 2.2.1 are representing the worst-case scenario. The real power of these approximation is in their average accuracy [for instance see Eq. (B.26)]. This is because during optimization process the average behavior of the approximation over the parameters of interest matters the most. It should be also noted that the only advantage of $MTF^{a2}$ over $MTF^{a1}$ is its simple expression. If one is interested in numerical rather than analytical

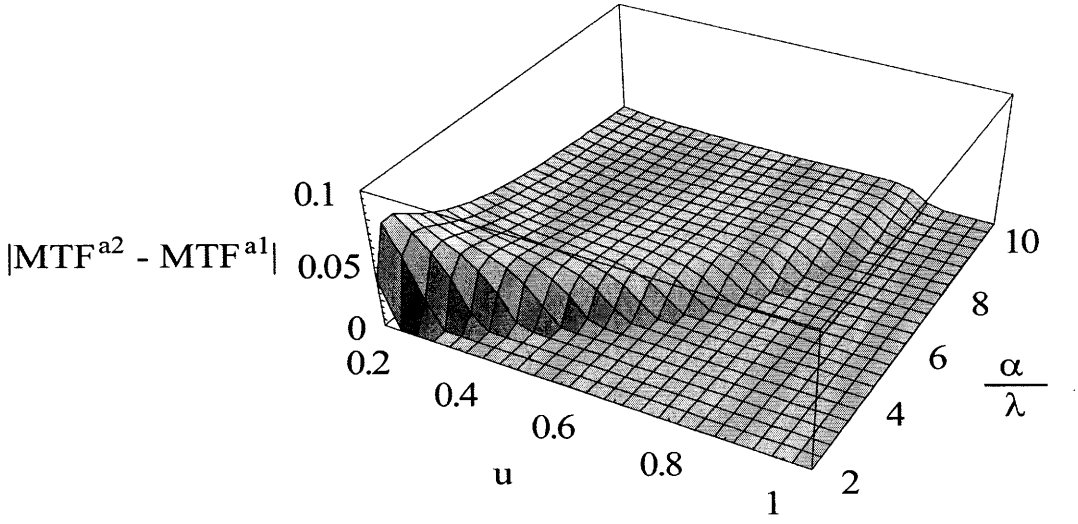



Figure 2-3: Plot of $\left|MTF^{a2} - MTF^{a1}\right|$ as a function of $\hat{\alpha}$ and $u$ when $\hat{W}_{20} = 4$ (Note that all variables are dimensionless).

optimization, then $MTF^{a1}$ is the approximation expression to be chosen. This is because it has a good accuracy and a closed form expression and hence much easier to compute compared to $MTF^{e}$ which involves the calculation of a double integral.

Finally it is worth mentioning that although these MTF expressions only consider defocus and cubic phase shift, one can employ the same approach to incorporate other aberrations into the approximate MTF expressions. The key is to capture the mean behavior of the integrand and then do the integral. This way, one first keeps the accuracy high and second makes the job of evaluating the integral simple. To expand these MTF expressions to other aberrations is one of the possible future work directions.

In the next two Sections we show how by using this expression we can solve two classes of design problems that illustrate well the use of the cubic phase pupil function to extend the depth of field of an imaging system. In the first problem we discuss the *uniform image quality* imaging and in the second problem we discuss the *task-based* imaging.

## 2.3 Optimization for Uniform Image Quality

### 2.3.1 Statement of the Problem

A well-known challenge in imaging systems is how to maintain a uniform image quality as the object moves along the depth of field of interest [21, 18]. The term *image quality* is clearly very broad and, in general, the image quality not only depends on the imaging system but also on the object spectrum. In this Chapter, by *image quality* of an imaging system, we precisely mean two factors: (i) the range of spatial frequencies that can be transfered from the object to the image plane and (ii) the transfer function of the imaging system for this range of spatial frequencies. Thus a *uniform image quality* imaging system has (i) a fixed range of transferable spatial frequencies as the object moves through the desired range of depth of field and (ii) a uniform transfer function for that range of spatial frequencies as the object moves through the desired range of depth of field.

The design goal in this Section is to increase this uniform image quality. We assume that the range of the spatial frequencies of interest is given as a design specification and we try to maximize the transfer function of the imaging system for that range of spatial frequencies as the object moves through the desired range of depth of field. This is what we mostly encounter in typical photography. In this type of imaging systems, one is interested in getting a high quality image for some desired range of depths of field. Another example in typical photography is when we focus on a particular object and we would like to have a uniform quality over all parts of the image; i.e. not only the part we have focused on, but also in all other parts of the picture. This means we want to have a uniform image quality as we are moving through some depths of field of interest [30]. This is further clarified in Fig. 2-4. This figure shows two pictures taken from a non-planar object. The object is a stack of printer cartridges. They are positioned so that the leftmost is closest to the imaging system and the rightmost is furthest from the imaging system. Part (a) is the picture taken by a traditional imaging system. As it can be seen the image quality is not uniform in this picture. Part (b) is the picture taken by an imaging system which uses pupil function engineering. As it can be seen, this picture has a uniform quality over the range of depth of field of interest. In other words,

(a) (b)

Figure 2-4: Photos of a non-planer object with (a) traditional imaging system and (b) imaging system with pupil function engineering. Part (a) does not have a uniform image quality. Quality is good at best focus in the center of the photo and image gets blurry out of focus. Part (b) has a uniform image quality; i.e. the image quality is the same both in and out of focus. Here by uniform image quality we mean the uniform transfer function of the imaging system at the spatial frequencies of interest over the depth of field.

in part (b) the transfer function for the range of spatial frequencies of interest is uniform over the depth of field of interest, and thus the image quality is preserved in the depth of field of interest.

In this context the design goal can be stated as maximizing the entire MTF (which is a measure of the transfer function of the imaging system as a function of spatial frequencies) over the imaging volume. The optimization problem will be defined in a way to reach this goal.

Considering the preceding imaging design problem, the typical problem specifications are: range of object distances (range of $d_o$), focal length ($f$), aperture diameter ($D$), and maximum spatial frequency of interest for the *image* ($S_{fi}$). Using these fixed problem specifications and through the optimization process we find the design parameters: cubic phase coefficient ($\alpha$) and image-plane to exit-pupil distance ($d_i$). Our goal is to find the design parameters that satisfy the following optimality criterion

$$\max_{\alpha,d_i}\left\{\min_{d_o,u}\left\{MTF^{a2}(u,0)\right\}\right\}. \tag{2.5}$$

$$\alpha \in \Re$$

$$d_i \in \Re$$

$$d_o \in [d_{o1}, d_{o2}]$$

$$u \in [0, u_{max}]$$

Equation (2.5) along with Eq. (2.4) are used as the basis of the optimization in Section 2.3.2. The analytic expression for design parameters ($\alpha$ and $d_i$) are found as a solution of the optimization problem.

### 2.3.2 Optimization

In this Section we solve the optimization problem stated in Eq. (2.5). We begin with a discussion about the highest normalized spatial frequency in the image plane, $u_{max}$. Since our aim is to have *uniform image quality*, $u_{max}$ is constant over the range of optimization parameters. Its value is defined using the highest spatial frequency of interest for the image ($S_{fi}$) as below

$$u_{max} = \frac{S_{fi}}{2f_o}, \tag{2.6}$$

where $f_o$ is the diffraction limited spatial frequency of the coherent imaging system. Thus we have

$$u_{max} = \frac{2\pi d_i S_{fi}}{kD}. \tag{2.7}$$

Now, considering Eq. (2.4) and using the fact that $MTF^{a2}(u,0)$ is monotonically decreasing when $u \in [0,1]$ (see Appendix E for a rigorous proof), we can conclude that minimization over $u$ is

equivalent to setting $u = u_{max}$. Hence, we can rewrite Eq. (2.5) as

$$\max_{\alpha,d_i}\left\{\min_{d_o}\left\{MTF^{a2}\left(\frac{2\pi d_i S_{fi}}{kD},0\right)\right\}\right\}. \tag{2.8}$$
$$\alpha \in \Re$$
$$d_i \in \Re$$
$$d_o \in [d_{o1}, d_{o2}]$$

The next step is the minimization of the $MTF^{a2}$ over $d_o$ and the maximization of the $MTF^{a2}$ over $d_i$. In fact these two steps are coupled, for they both have a direct effect on $W_{20}$, and thus on the system's defocus, as it is explicitly shown in Eq. (A.6). In this step, we use the fact that increasing the absolute defocus ($W_{20}$) reduces $MTF^{a2}$ and vice versa (see Appendix E for details). Since increasing the absolute defocus reduces $MTF^{a2}$ and vice versa, one can define a sub-optimization problem for these two parameters as shown below

$$\min_{d_i}\left\{\max_{d_o}\{|W_{20}|\}\right\}. \tag{2.9}$$
$$d_i \in \Re$$
$$d_o \in [d_{o1}, d_{o2}]$$

Recall that $W_{20}$ is given by Eq. (A.6). Using elementary calculus, one can solve the problem above to find

$$d_i^* = \frac{2f d_{o1} d_{o2}}{2d_{o1}d_{o2} - f(d_{o1}+d_{o2})}, \tag{2.10}$$

$$d_o = d_{o1} \text{ or } d_{o2}.$$

Thus we can rewrite Eq. (2.8) as

$$\max_{\alpha} \left\{ MTF^{a2}\left(u_{max}, 0\right) \right\}, \qquad \alpha \in \Re \tag{2.11}$$

where $u_{max}$ is defined through Eqs. (2.7) and (2.10) and $W_{20}$ is defined through Eqs. (A.6) and (2.10). Note that neither $u_{max}$ nor $W_{20}$ is a function of $\alpha$ and thus one can easily deal with this optimization problem without worrying about the complicated formulas for $u_{max}$ and $W_{20}$ (As we will see, this is not the case in Section 2.4).

To solve Eq. (2.11) we find the maximum value of the $MTF^{a2}$ by setting its first derivative equal to zero

$$\frac{\partial MTF^{a2}(\alpha)}{\partial \alpha} = 0. \tag{2.12}$$

Note that this approach is only feasible because we are using an approximation to the exact MTF. In fact, the exact MTF is a highly oscillating function, and such an approach for finding the optimal cubic phase coefficient ($\alpha^*$) is of little use. Strictly speaking, in case of the exact MTF, Eq. (2.12) does not have a unique solution in the regions of interest in imaging design. However, our approximation ($MTF^{a2}$) which represents the average value of this oscillating function ($MTF^{e}$) results in a unique optimal value for the cubic phase coefficient ($\alpha^*$) which could be found through Eq. (2.12) (See Appendix E for more details about the properties of our approximation). This fact allows solving Eq. (2.12) for the optimal cubic phase coefficient ($\alpha^*$). An approximate solution to Eq. (2.12) is found in Appendix D. The result is as follows

$$\frac{\alpha^*}{\lambda} = \frac{1 + 8u_{max}\frac{W_{20}}{\lambda}(1-u_{max}) + \sqrt{1 + 16u_{max}\frac{W_{20}}{\lambda}(1-u_{max})}}{24u_{max}(1-u_{max})^2}, \tag{2.13}$$

where $u_{max}$ is defined through Eqs. (2.7) and (2.10) and $W_{20}$ is defined through Eqs. (A.6) and (2.10). In Section 2.3.3 we will discuss the optimization results obtained in this Section.

### 2.3.3 Results

In this Section we discuss the results of the optimization which was done in the last Section. We use a sample problem to clarify the benefit of using this method. We also present the graphs of the general results along with a method of how to use these graphs in a practical optical design problem.

We begin with presenting the final results of optimization in Eqs. (2.14). As it could be seen through these equations, all the design parameters are expressed in terms of the problem specifications; i.e. $f$, $D$, $k$, $d_{o1}$, $d_{o2}$ and $S_{fi}$.

$$\begin{aligned} \frac{\alpha^*}{\lambda} &= \frac{1+8u_{max}\frac{W_{20}}{\lambda}(1-u_{max})+\sqrt{1+16u_{max}\frac{W_{20}}{\lambda}(1-u_{max})}}{24u_{max}(1-u_{max})^2}, \\ d_i^* &= \frac{2fd_{o1}d_{o2}}{2d_{o1}d_{o2}-f(d_{o1}+d_{o2})}, \end{aligned} \tag{2.14}$$

where $u_{max}$ and $W_{20}$ are

$$\begin{aligned} u_{max} &= \frac{2\pi d_i S_{fi}}{kD}, \\ W_{20} &= \frac{D^2}{8}\left(\frac{1}{d_i}+\frac{1}{d_{o1}}-\frac{1}{f}\right). \end{aligned} \tag{2.15}$$

In order to illustrate the results of the optimization, we use a sample imaging design problem, whose specifications are shown in Table 2.1. The wave number, $k = 2\pi n/\lambda$, is chosen to be the average value for visible light propagating in air. The aperture diameter, $D$, and the focal length, $f$ are chosen so that we have a practically feasible $f\#$ at a reasonable cost. The required depth of field, i.e. $d_{o1}$ and $d_{o2}$, which should be chosen according to the goal for the range of functioning of imaging

Table 2.1: Problem specifications for the sample uniform image quality imaging problem.

| Param. | Value | Unit |
|---|---|---|
| $k$ | $11.4 \times 10^{+3}$ | $mm^{-1}$ |
| $D$ | 8 | $mm$ |
| $f$ | 50 | $mm$ |
| $d_{o1}$ | 450 | $mm$ |
| $d_{o2}$ | 710 | $mm$ |
| $S_{fi}$ | 90 | $\frac{line-pair}{mm}$ |

Table 2.2: Optimized designed parameters for uniform image quality imaging problem.

| Param. | Value | Unit |
|---|---|---|
| $\frac{\alpha^*}{\lambda}$ | 4.60 | — |
| $d_i^*$ | 55 | $mm$ |
| $u_{max}$ | 0.34 | — |
| $\frac{W_{20}}{\lambda}$ | 6 | — |

system, is chosen to be in accordance with its corresponding value in the sample task-based problem so that one could compare results obtained for each method. The maximum spatial frequency of interest in the image plane, $S_{fi}$, is chosen with the same goal in mind.

Using the problem specification values presented in Table 2.1, one can get the optimized design variables using Eqs. (2.14) and (2.15). These values are shown in Table 2.2. To compare the performance of the optimized imaging system with that of the traditional system (without pupil function engineering: $d_i = 54.7mm$ and $\alpha = 0$), we have shown the plot of the $MTF^e$ at three different depths of field for these two systems in Fig. 2-5. Using this figure, one can see how the $MTF^e$ has been distributed over the entire field rather than just at the in-focus plane. One can also see that the worst-case MTF (minimum $MTF^e$ in the union of the range of interest of all variables) is 0.10. Also, the graph of defocus, $W_{20}$, versus depth of field is shown in Fig. 2-6. This graph shows that we have actually minimized the maximum absolute value of defocus along the desired depth of field; i.e. we have reduced $W_{20}$ from $7\lambda$ to $6\lambda$. Note that, in the optimized system, the best focus has moved from the middle of depth of field toward the lens. In particular $d_o$ for best focus is reduced from $0.58\,m$ to $0.55\,m$. This is due to nonlinear relationship between $W_{20}$ and $d_o$.

For any other uniform quality imaging problem, one will only need to use the problem specifi-

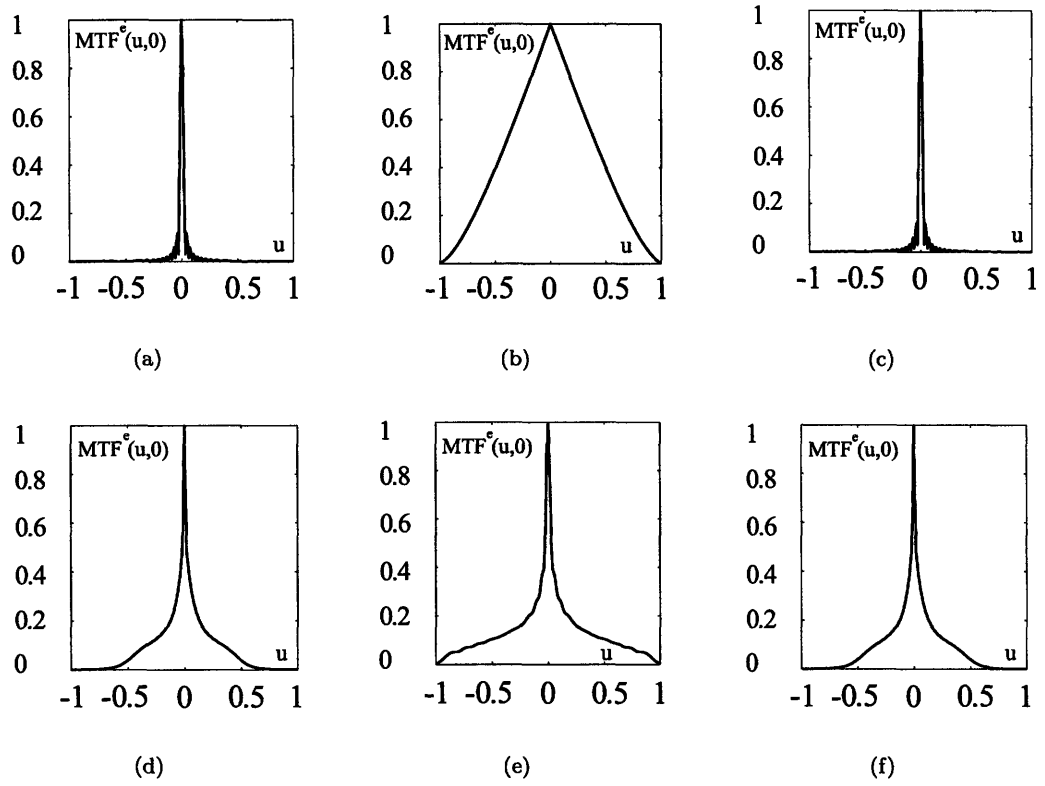



Figure 2-5: $MTF^e(u,0)$ of the system with and without pupil function engineering (Uniform image quality imaging problem; optical system specifications are from Tables 2.1 and 2.2). Note how image quality (the transfer function of the imaging system at the spatial frequencies of interest) is uniform over the depth of field. (a) Traditional system (far field, $\frac{W_{20}}{\lambda} = -5$). (b) Traditional system (in focus, $\frac{W_{20}}{\lambda} = 0$). (c) Traditional system (near field, $\frac{W_{20}}{\lambda} = +7$). (d) Optimized system (far field, $\frac{W_{20}}{\lambda} = -6$). (e) Optimized system (in focus, $\frac{W_{20}}{\lambda} = 0$). (f) Optimized system (near field, $\frac{W_{20}}{\lambda} = +6$).

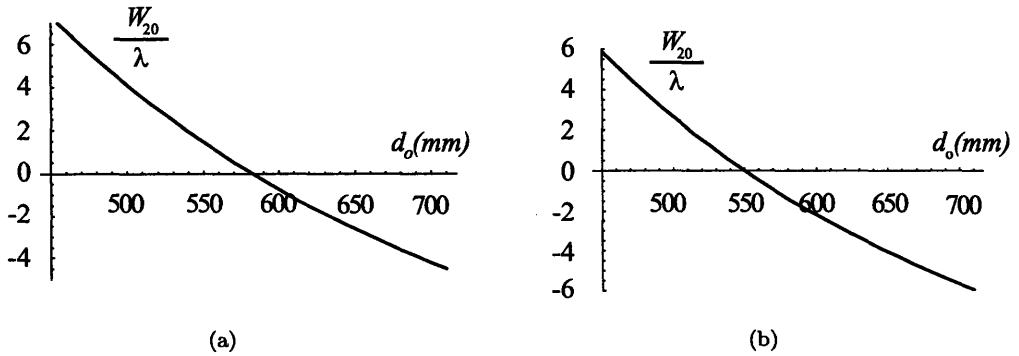



Figure 2-6: Defocus of the (a) traditional imaging system and (b) Optimized imaging system in uniform quality imaging problem (Optical system specifications are from Tables 2.1 and 2.2). Note that in the optimized imaging system the best focus has been moved toward the lens to reduce the maximum absolute defocus from $7\lambda$ to $5.5\lambda$.

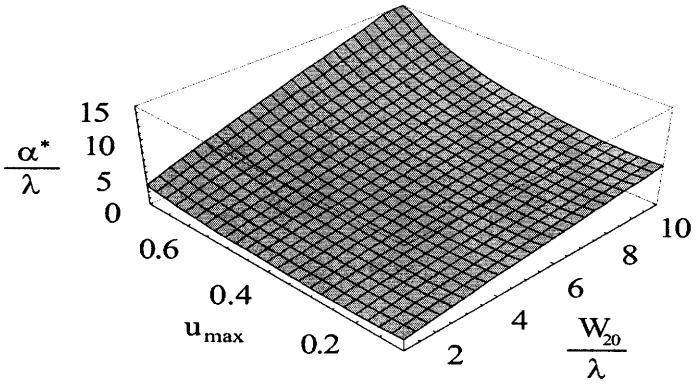


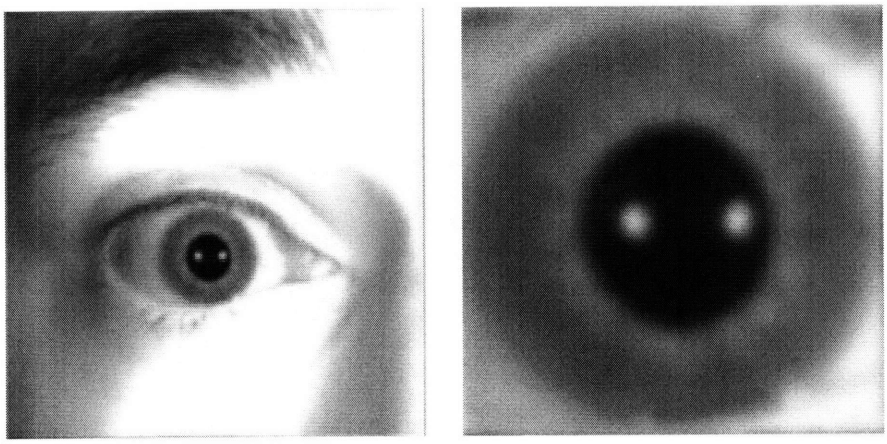

Figure 2-7: Optimum cubic phase coefficient ($\alpha^*$) for the uniform image quality problem. Using given problem specifications one can find the corresponding $u_{max}$ and $W_{20}$ (Eq. (2.15)), and then $\alpha^*$ can be directly read from this figure [Eq. (2.14)].

cations and evaluate $W_{20}$ and $u_{max}$ through Eqs. (2.15) and then use these values for evaluating $d_i^*$ and $\alpha^*$ through Eqs. (2.14). Figure 2-7 shows the optimal value of the cubic phase coefficient versus defocus and image quality (maximum spatial frequency of interest in the image plane).

## 2.4 Optimization for Task-based Imaging

### 2.4.1 Statement of the Problem

Task-based imaging systems have played an important role in the development of industrial applications and in the improvement of living standards in recent years. These roles range from simple bar code reading in a supermarket to complicated identification systems in high-security facilities (for instance, biometric iris recognition [20]). A critical challenge in this field is to have a sufficiently good image in a certain depth of field, in the sense that this image must be usable for the specific task. Particularly, for task-based imaging it is immaterial whether the picture looks good or not. Rather, the amount of usable information that is transfered from the target object to our detective device is of utmost importance. In general, this means we need more resolution when the image is smaller and less when the image is bigger. This is clarified further in Fig. 2-8, which shows two photos are taken by an iris recognition system. As it can be expected, the system is only concerned



(a) (b)

Figure 2-8: Iris recognition images as an example of task-based imaging. (a) Far field image ($d_o = 800mm$). (b) Near field image ($d_o = 200mm$). Although part (a) appears to be a higher quality image, parts (a) and (b) both have equal usable information of iris.

about the amount of information that it captures from the iris. Thus, as the person gets closer to the device, the required resolution decreases so that the amount of information from the iris remains constant. Although part (a) may be considered a good picture and part (b) a poor one from the photography point of view, they are both equally good from a task-based point of view.

In the design process of a task-based imaging system, it is crucial to take the above point into account. It is often the case that we want to capture a constant amount of usable information from the object as the object is moving along a desired depth of field. This instantly calls for the use of pupil function engineering. However, in this case the optimal design criterion would be different from the one we discussed in the last Section.

In an imaging problem like this, the classic problem specifications are depth of field of interest, i.e. the range along which the object can move (range of $d_o$), focal length ($f$), aperture diameter ($D$), and maximum spatial frequency of interest for the *object*, i.e. the maximum amount of detail from the object that we need to capture for our specific task ($S_{fo}$). Using these problem specifications and through the optimization process we find the design parameters, which are the cubic phase coefficient ($\alpha$) and the image-plane to exit-pupil distance ($d_i$). Our goal is to find design parameters that satisfy the optimality criterion. This is expressed in Eq. (2.16):

$$\max_{\alpha,d_i}\left\{\min_{d_o,u}\left\{MTF^{a2}(u,0)\right\}\right\}. \tag{2.16}$$

$$\alpha \in \Re$$

$$d_i \in \Re$$

$$d_o \in [d_{o1}, d_{o2}]$$

$$u \in [0, u_{max}(d_o)]$$

Equation (2.16) along with Eq. (2.4) are used as the basis of the optimization in the next Section. Note the fundamental difference between Eq. (2.16) and (2.5); in Eq. (2.5) the maximum normalized spatial frequency of interest in the image plane ($u_{max}$) is constant whereas in Eq. (2.16) it is a function of $d_o$ and it changes as the object moves along the desired range of depths of field. This results in a coupled optimization problem that is clearly more complex than the one solved in Section 2.3. In the optimization process shown in the next Section, the analytic expression for the design parameters ($\alpha$ and $d_i$) are found.

### 2.4.2 Optimization

We begin with a discussion about the normalized spatial frequency of interest in the image plane, $u_{max}$. Using its definition, we have

$$u_{max} = \frac{S_{fi}}{2f_o}, \tag{2.17}$$

where $f_o$ is the diffraction limited spatial frequency of the coherent system. Since we want to have constant usable information transfer from the object to our detective device, the maximum spatial frequency of interest for the image ($S_{fi}$) changes as the object moves along the range of interest of depths of field. In fact, $S_{fi}$ is simply related to $S_{fo}$ through the following relation

$$S_{fi} = \frac{S_{fo}}{M}, \tag{2.18}$$

where $M$ is the lateral magnification. Replacing the lateral magnification, $M$, and the diffraction limit spatial frequency, $f_o$, with their corresponding values, we have

$$u_{max} = \frac{2\pi S_{fo} d_o}{kD}. \tag{2.19}$$

Now, using the same line of reasoning as in Section 2.3, Eq. (2.16) can be reduced to

$$\max_{\alpha, d_i} \left\{ \min_{d_o} \left\{ MTF^{a2} \left( \frac{2\pi S_{fo} d_o}{kD}, 0 \right) \right\} \right\}. \tag{2.20}$$

$$\alpha \in \Re$$

$$d_i \in \Re$$

$$d_o \in [d_{o1}, d_{o2}]$$

Unlike the last Section, where $d_o$ was present in the equation of $MTF^{a2}$ [Eq. (2.4)] only as a part of $W_{20}$, here it also appears in the expression of $u_{max}$. Hence we cannot set up a sub-optimization problem to just maximize $W_{20}$ over $d_o$ and find $d_o$. To overcome this complexity let us define the partial defocus, $W_{20}^{'}$ as

$$W_{20}^{'} = \frac{D^2}{8} \left( \frac{1}{d_i} - \frac{1}{f} \right). \tag{2.21}$$

Note that $W_{20}^{'}$ contains all the system defocus terms, except for the $1/d_o$ term. Defining partial defocus as above helps us keep track of the effect of changing $d_o$ on $MTF^{a2}$ independently from the

rest of defocus terms. Using properties of our approximation (see Appendix E) the optimization problem over $d_o$ and $d_i$ in Eq. (2.20) can be rewritten as below

$$MTF^{a2}\left(\frac{2\pi S_{fo} d_{o1}}{kD}, 0\right) = MTF^{a2}\left(\frac{2\pi S_{fo} d_{o2}}{kD}, 0\right). \tag{2.22}$$

To have more intuition about why this is the case, observe the $MTF^{a2}\left(\frac{2\pi S_{fo} d_o}{kD}, 0\right)$ in Fig. 2-9 where its graph is shown as a function of $d_o$ and $W_{20}^{'}$ [with $k\alpha = 10$, $kD^2/8 = 10^4$, $2\pi S_{fo}/(kD) = 10^{-3}$]. From Fig. 2-9 one can observe that for any range of interest of $d_o$, there is a particular value of $W_{20}^{'}$ which is optimal according to Eq. (2.20). It is also clear from this figure, that the corresponding optimal value of $d_o$ is either of $d_{o1}$ or $d_{o2}$. This is because $MTF^{a2}\left(\frac{2\pi S_{fo} d_o}{kD}, 0\right)$ monotonically decreases on each of its branches as $d_o$ approaches the corresponding end limit. This leads us to *maximize* $MTF^{a2}\left(\frac{2\pi S_{fo} d_o}{kD}, 0\right)$ over $W_{20}^{'}$ (which is equivalent to maximizing over $d_i$), and *minimize* $MTF^{a2}\left(\frac{2\pi S_{fo} d_o}{kD}, 0\right)$ over $d_o$, by choosing the particular value for $W_{20}^{'}$ suggested in Eq. (2.22).

Thus the sub-optimization problem in this Section is reduced to finding $W_{20}^{'}$ such that Eq. (2.22) is satisfied. This can be seen with the aid of two dashed lines in Fig. 2-9. Note that once the right amount of $W_{20}^{'}$ is found, then one can increase the MTF in that region using the optimization over $\alpha$. Solving Eq. (2.22) for $W_{20}^{'}$, we have

$$\frac{W_{20}^{'\,*}}{\lambda} = \frac{48\pi \frac{\alpha}{\lambda} d_{o1} d_{o2} (u_{max2} - u_{max1}) - kD^2 (d_{o1} + d_{o2})}{32\pi d_{o1} d_{o2}} \tag{2.23}$$

Note that $u_{max1}$ and $u_{max2}$ are defined through Eq. (2.19) after replacing $d_o$ by $d_{o1}$ and $d_{o2}$, respectively. Now the optimization problem can be rewritten as

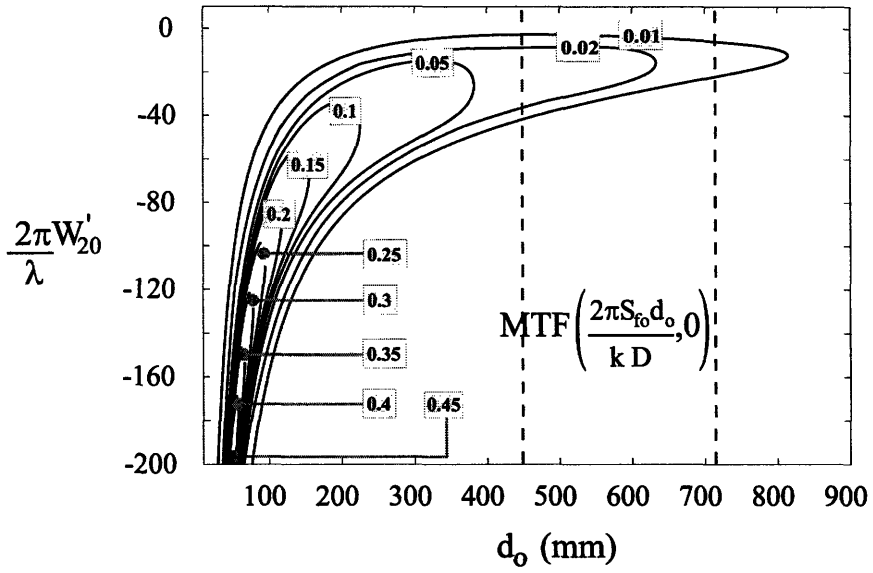



Figure 2-9: Behavior of the $MTF^e$ with respect to partial defocus ($W'_{20}$) and depth of field ($d_o$). The region between the dashed lines represents the depth of field of interest. The goal is to have maximum $MTF^e(u_{max}(d_o), 0) = MTF^e\left(\frac{2\pi S_{fo} d_o}{kD}, 0\right)$ in this region. To do so we find the $W'_{20}$ for which $MTF^{a2}\left(\frac{2\pi S_{fo} d_o}{kD}, 0\right)$ is the same at both ends of this region of interest [see Eq. (2.22)]. This is justified by assuming that $MTF^{a2}$ has a parabolic behavior with respect to $W'_{20}$. In this figure we have used $k\alpha = 10$, $kD^2/8 = 10^4 mm$ and $2\pi S_{fo}/(kD) = 10^{-3} mm^{-1}$.

$$\max_{\alpha} \left\{ MTF_w^{a2} \left( \frac{2\pi S_{fo} d_{o1}}{kD}, 0 \right) \right\}, \tag{2.24}$$
$$\alpha \in \Re$$

where $MTF_w^{a2}$ in Eq. (2.24) contains the optimal value of partial defocus, $W_{20}^{'*}$, as in Eq. (2.23). Since optimal partial defocus is also a function of $\alpha$, one cannot use the result of optimization in Section 2.3. This is because the structure of $MTF_w^{a2}$ with respect to $\alpha$ is different from that of $MTF^{a2}$. This calls for a more involved optimization problem to be dealt with. To solve Eq. (2.24) we find the maximum value of the $MTF_w^{a2}$ by setting its first derivative equal to zero

$$\frac{\partial MTF_w^{a2}(\alpha)}{\partial \alpha} = 0. \tag{2.25}$$

Eq. (2.25) is solved using the same method as Eq. (2.12) (Which is explained in Appendix D), except that now, the starting point is

$$\begin{aligned} MTF_w^{a2}(\alpha) \;=\; & \frac{c_1}{\alpha} + \frac{c_2}{\sqrt{\alpha}} \times \\ & \left[ \text{Arctan}\left(\frac{c_3}{\sqrt{\alpha}} + c_4\sqrt{\alpha} + c_5\sqrt{\alpha}\right) - \text{Arctan}\left(\frac{c_3}{\sqrt{\alpha}} - c_4\sqrt{\alpha} + c_5\sqrt{\alpha}\right) \right], \end{aligned} \tag{2.26}$$

rather than Eq. eC1??. Solving Eq. (2.25), we have

$$\alpha^* = \frac{1 + 2c_3c_4 + c_3c_5 + \sqrt{1 + 4c_3c_4 + 4c_3^2c_5^2}}{2\left(c_5^2 - c_4^2\right)}, \tag{2.27}$$

or

$$\frac{\alpha^*}{\lambda} = \frac{u_2 + 2C\Delta u(\Delta u + 1 - u_1) + \sqrt{u_2^2 + 4Cu_2\Delta u^2 + 16C^2(1-u_1)^2\Delta u^2}}{12\pi u_1 u_2(2 - u_1 - u_2)\left[\Delta u + 2(1-u_1)\right]}, \tag{2.28}$$

where $u_1$ and $u_2$ are defined through Eq. (2.19) after replacing $d_o$ by $d_{o1}$ and $d_{o2}$. $C$ and $\Delta u$ are defined as

$$\begin{aligned} C &= \frac{\pi S_{fo} D}{4}, \\ \Delta u &= u_2 - u_1. \end{aligned} \tag{2.29}$$

### 2.4.3 Results

We begin this Section by presenting the final results of optimization in Eqs. (2.30). As it can be seen through these equations, all the design parameters are expressed in terms of the problem specifications; i.e. $f$, $D$, $k$, $d_{o1}$, $d_{o2}$ and $S_{fo}$.

$$\begin{aligned} \frac{\alpha^*}{\lambda} &= \frac{u_2 + 2C\Delta u(\Delta u + 1 - u_1) + \sqrt{u_2^2 + 4Cu_2\Delta u^2 + 16C^2(1-u_1)^2\Delta u^2}}{12\pi u_1 u_2(2 - u_1 - u_2)\left[\Delta u + 2(1-u_1)\right]}, \\ \frac{1}{d_i{}^*} &= \frac{1}{f} + \frac{48\pi \frac{\alpha^*}{\lambda} d_{o1}\, d_{o2}\,(u_2 - u_1) - kD^2\,(d_{o1} + d_{o2})}{2d_{o1}d_{o2}kD^2}, \end{aligned} \tag{2.30}$$

where $u_1$, $u_2$, $\Delta u$ and $C$ are defined as

Table 2.3: Problem specifications for the sample task-based imaging problem.

| Param. | Value | Unit |
|---|---|---|
| $k$ | $11.4 \times 10^{+3}$ | $mm^{-1}$ |
| $D$ | 8 | $mm$ |
| $f$ | 50 | $mm$ |
| $d_{o1}$ | 450 | $mm$ |
| $d_{o2}$ | 710 | $mm$ |
| $S_{fo}$ | 7 | $\frac{line-pair}{mm}$ |

$$
\begin{aligned}
u_1 &= \frac{2\pi S_{fo} d_{o1}}{kD}, \\
u_2 &= \frac{2\pi S_{fo} d_{o2}}{kD}, \\
\Delta u &= u_2 - u_1, \\
C &= \frac{\pi S_{fo} D}{4}.
\end{aligned}
\tag{2.31}
$$

In order to illustrate the results of optimization, we use a sample problem. We consider an iris recognition system as an example of task-based imaging. The problem specifications are shown in Table 2.3. Although the wave number, $k$, typically used for iris recognition is that of *near-infra-red*, for the sake of comparison of results with the case of photography (Section 2.3), the wave number is chosen to be the average value of visible light. It should be noted that the method performs satisfactory for the near-infra-red wavelength as well. The aperture diameter, $D$, and the focal length, $f$ are chosen according to manufacturing and size limitations. The required depth of field, i.e. $d_{o1}$ and $d_{o2}$, are chosen to satisfy the goal of system operation without any need for cooperation from the user. The maximum spatial frequency of interest in the object plane, $S_{fo}$, is chosen to satisfy the minimum number of pixels across the iris, required by the recognition algorithm.

Using the problem specification values presented in Table 2.3, one can get the optimized design variables using Eqs. (2.30) and (2.31). These values are shown in Table 2.4.

To compare the performance of the optimized imaging systems, with that of the traditional system (without pupil function engineering: $d_i = 54.7mm$ and $\alpha = 0$), we have shown the plot of the

Table 2.4: Optimized designed parameters for the sample task-based imaging problem.

| Param. | Value | Unit |
|---|---|---|
| $\frac{\alpha^*}{\lambda}$ | 4.06 | — |
| $d_i^*$ | 54.85 | $mm$ |
| $u_{max1}$ | 0.22 | — |
| $u_{max2}$ | 0.34 | — |
| $\frac{W_{20}^{'*}}{\lambda}$ | −25.64 | — |

$MTF^e$ at three different depths of field for these two systems in Fig. 2-10. Using this figure, we can see how regions of spatial frequency with high value of $MTF^e$ have been distributed in an optimal way among all the depths of field rather than just at best focus. The vertical dashed lines are the lines with spatial frequency $u = u_{max}(d_o)$; i.e. they represent the maximum spatial frequency of interest in that particular depth of field. It could be read from this figure that the worst-case MTF (minimum $MTF^e$ in the union of the range of interest of all variables) is 0.11.

Figure 2-11 shows the graph of $MTF^e\left(\frac{2\pi S_{fo} d_o}{kD}, 0\right)$ as a function of partial defocus and depth of field for the optimized system (using the system parameters given in Table 2.4). While the optimization over partial defocus, $W_{20}^{'}$, has kept the minimum MTF at both ends of the desired range of the depth of field equal, the optimization over $\alpha$ has increase MTF at the prescribed value of partial defocus.

Also in Fig. 2-12, the graph of defocus, $W_{20}$, versus depth of field is shown. This graph shows how we have actually increased the absolute value of defocus compared to the traditional imaging system. This increased defocus is responsible for low quality images in the near field. However, as explained before, this is in fact an important advantage, because we are acquiring *just* the necessary usable information from the object (an iris in this example) and thus we are saving up modulation for the far field. Note that in the optimized system the best focus has moved closer to the far end of the depth of focus. We are thus freeing up modulation for the far field.

Note that the main difference of the uniform quality and task-based imaging problems is the sub-optimization that was done in the last Section on $W_{20}^{'}$. To further illustrate this sub-optimization, the graph of the $MTF^e(u_{max}(d_o), 0)$ versus $d_o$ is shown in Fig. 2-13. The solid line represents the

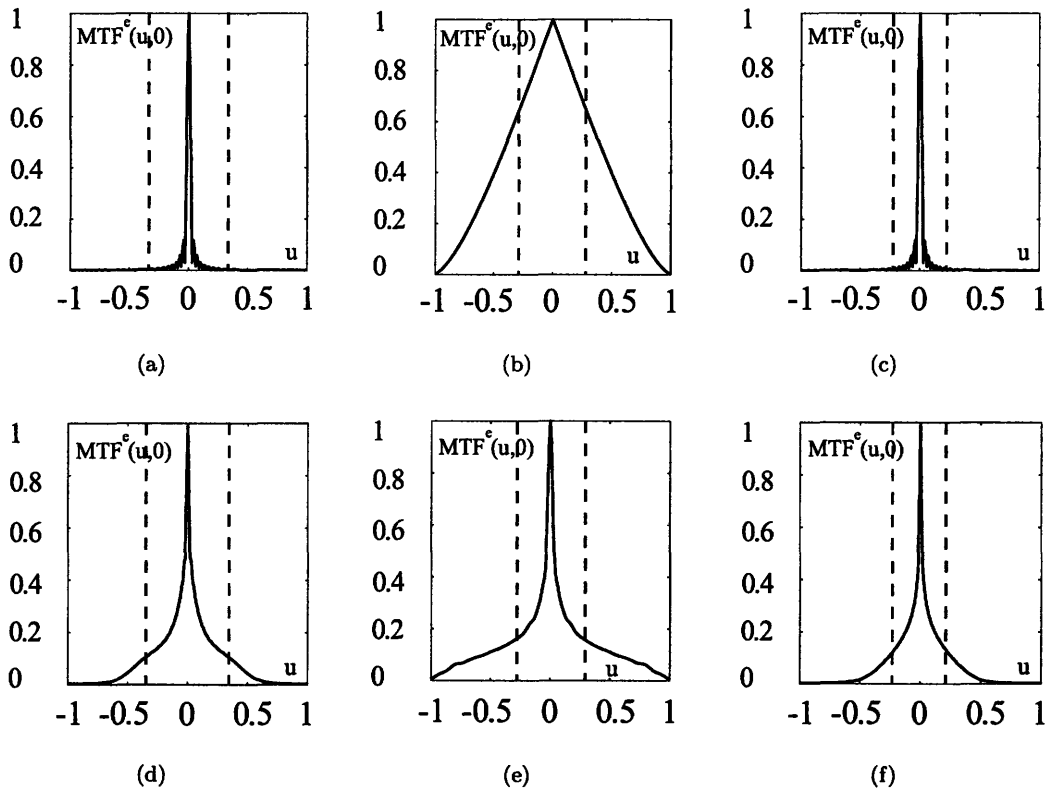



Figure 2-10: $MTF^e(u,0)$ of the system with and without pupil function engineering (Task-based imaging problem; optical system specifications are from Tables 2.3 and 2.4). The region between dashed lines represents the range of spatial frequencies of interest for that particular depth of field. Note how this range of spatial frequencies of interest gets smaller as the object gets closer to imaging system. (a) Traditional imaging system (far field, $\frac{W_{20}}{\lambda} = -5$), (b) Traditional imaging system (in focus, $\frac{W_{20}}{\lambda} = 0$), (c) Traditional imaging system (near field, $\frac{W_{20}}{\lambda} = +7$), (d) Optimized imaging system (far field, $\frac{W_{20}}{\lambda} = -4$), (e) Optimized imaging system (in focus, $\frac{W_{20}}{\lambda} = 0$), (f) Optimized imaging system (near field, $\frac{W_{20}}{\lambda} = +8$).

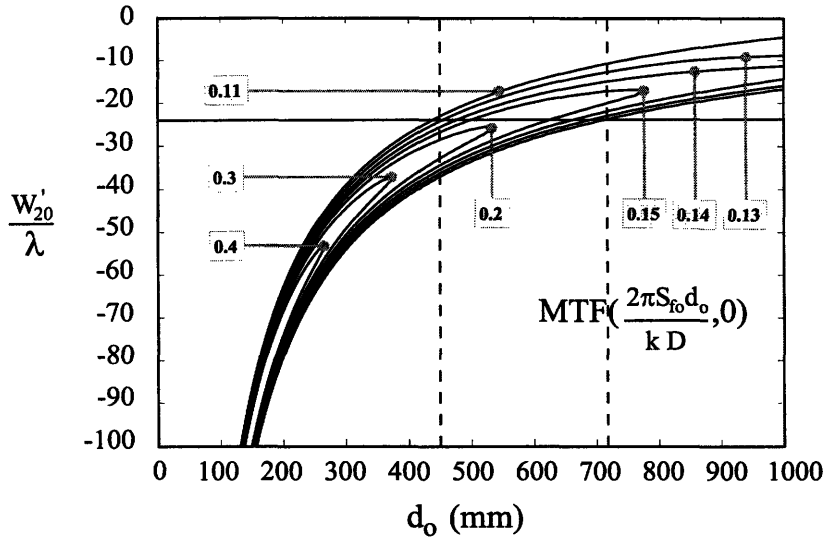



Figure 2-11: The $MTF^e$ as a function of partial defocus and depth of field for the optimum system (task-based imaging problem; optical system specifications are from Tables 2.3 and 2.4). The region between the dashed lines represents the depth of field of interest. The horizontal solid line represents the optimum value of $W_{20}^{'}$. As it can be seen the goal of maximizing the MTF is achieved. Note how $MTF^e\left(\frac{2\pi S_{fo} d_{o1}}{kD}, 0\right) \approx MTF^e\left(\frac{2\pi S_{fo} d_{o2}}{kD}, 0\right)$ as it is expected from Eq. (2.22).

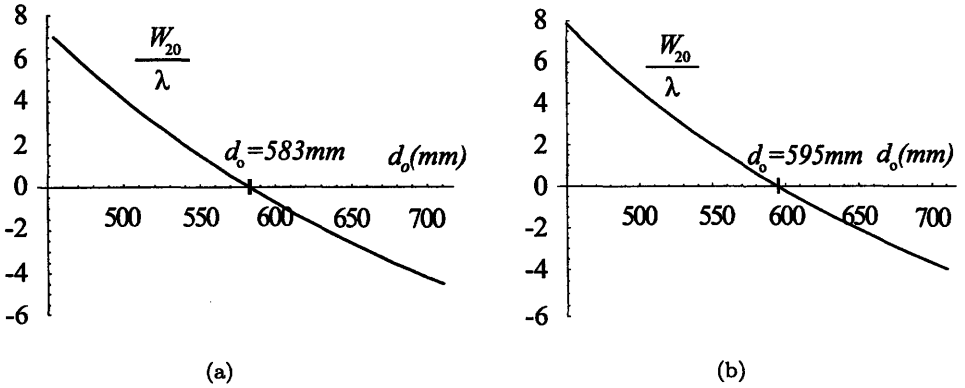



Figure 2-12: Defocus of the (a) traditional imaging system and (b) imaging system with optimized pupil function engineering (Task-based imaging problem; optical system specifications are from Tables 2.3 and 2.4). Note how in the optimized imaging system the best focus is moved far from the imaging system to balance the modulation at the highest spatial frequency of interest over the entire depth of field.

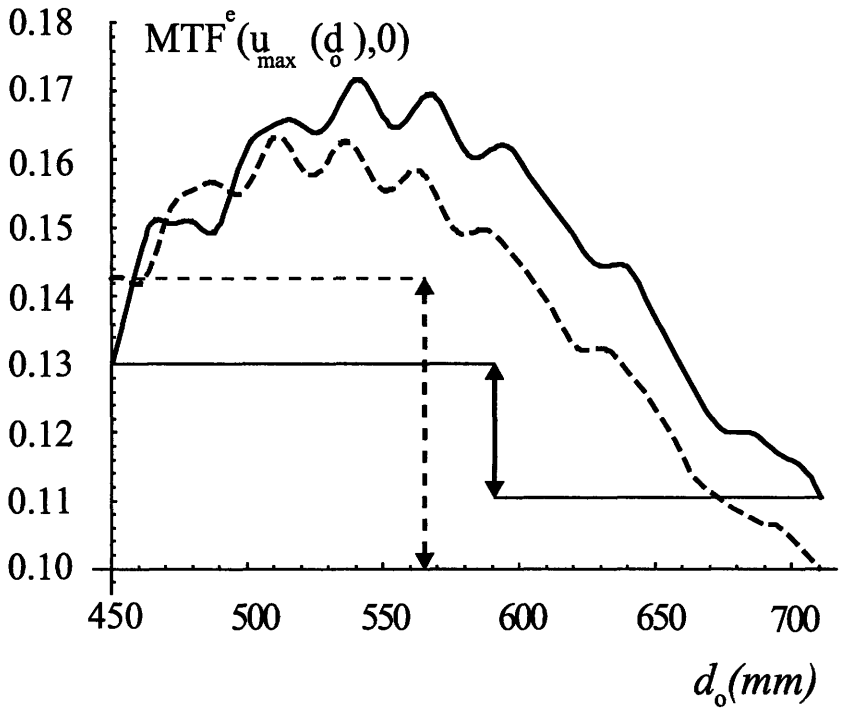



Figure 2-13: The minimum value of $MTF^e$ in the range of spatial frequencies of interest [namely $MTF^e(u_{max}(d_o),0) = MTF^e\left(\frac{2\pi S_{fo} d_o}{kD},0\right)$] v.s. depth of field. The solid line represents the optimized task-based imaging system and the dashed line represents the optimized uniform quality imaging system. This figure shows how the optimized uniform quality imaging system is not efficient for task specific imaging. Note how the sub-optimization of Eq. (2.22) has increased $MTF^e\left(\frac{2\pi S_{fo} d_o}{kD},0\right)$ over the depth of field of interest as shown by the solid-line graph. Optical system specifications are from Tables 2.2 and 2.4.

task-based optimized imaging system. The dashed line represents the uniform quality optimized imaging system. Clearly the goal of uniformly maximizing $MTF^e(u_{max}(d_o),0)$ over the range of interest of $d_o$ is achieved with the solid line in the figure.

For any other specific problem, one only needs to use the problem specifications and evaluate the optimal design parameters through Eqs. (2.30) and (2.31). Figures 2-14 and 2-15 show the graph of optimal parameters for the class of task-based imaging extension of depth of field problems using cubic phase element. Figure 2-14 shows the optimal value of the cubic phase coefficient versus the range of interest of the depth of field. One can observe that as depth of field gets larger the optimal value of the cubic phase coefficient gets larger too. Also note that this optimal value is symmetric with respect to change of $d_{o1}$ to $d_{o2}$. In Fig. 2-15 we have shown $D^2/(8d_i^*)$ versus range of interest

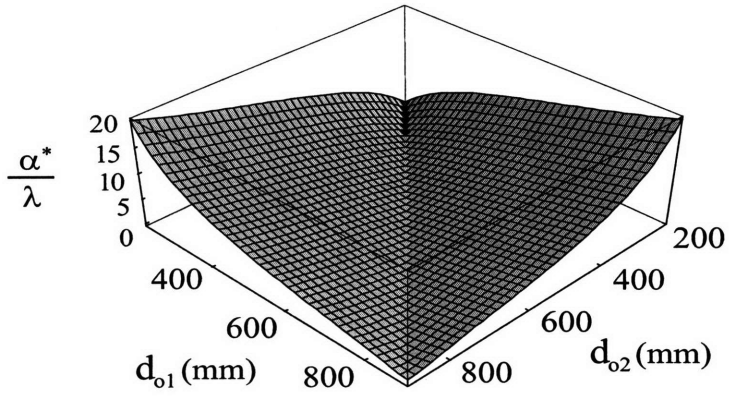



Figure 2-14: Optimum cubic phase coefficient ($\alpha^*$) for task-based imaging. Using the range of interest for object ($d_{o1}$ and $d_{o2}$), one can find the $\alpha^*$ from this figure [Eq. (2.30)]. In this figure we have used $\lambda = 0.55 \times 10^{-3} mm, D = 8mm$ and $S_{fo} = 14 \frac{line-pair}{mm}$.

of depth of field. To obtain the value of $d_i^*$, one may read the value of $D^2/(8d_i^*)$ from Fig. 2-15, and then evaluate $d_i^*$. Note that unlike $\alpha^*$, the optimum value of $D^2/(8d_i^*)$ is not symmetric with respect to change of $d_{o1}$ to $d_{o2}$. This is due to the asymmetric nature of optimization in this Section. This asymmetry forces the image quality not to be uniform as object moves along the desired depth of field. Rather, it tries to keep the amount of usable information transfered from object constant. Both of the above figures will have a tilt toward higher values of $\alpha^*$ and $D^2/(8d_i^*)$ as the object spatial frequency of interest, $S_{fo}$, increases.

## 2.5 Discussion

In general, two main categories of imaging systems are common photography, and task-based imaging. A common challenge in either of those is to increase the depth of field. In case of photography this increase results in a higher quality image for both the in-focus target and out-of-focus surroundings. The increase in the depth of field in photography is particularly important when we are dealing with a multi-target image where the targets are at different distances from the imaging system. In the case of task-based imaging, the robustness of the system (how well the system performs if the target is slightly out of focus) is often directly proportional to the imaging volume; i.e. the depth

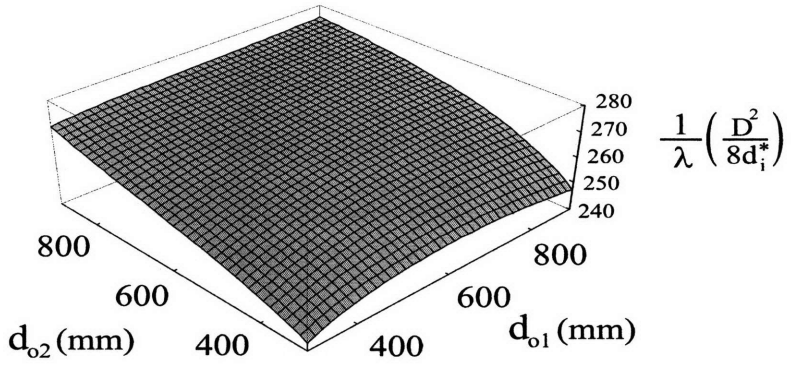



Figure 2-15: Optimum image-plane and exit pupil distance ($d_i^*$) for task-based imaging. Using the range of interest for object ($d_{o1}$ and $d_{o2}$), one can find the $d_i^*$ from this figure (Eq. (2.30)). In this figure we have used $\lambda = 0.55 \times 10^{-3} mm, D = 8mm, f = 50mm$ and $S_{fo} = 14 \frac{line-pair}{mm}$.

of field of the imaging system. Thus increasing the depth of field is a critical factor. Pupil function engineering is capable of increasing the depth of field of an imaging system. However, optimal design of the right pupil function engineering element for a particular system is often a challenge. In this Chapter, for the first time, we have solved this design problem analytically for the particular case of a cubic phase element. Equations (2.14) and (2.15) provide us with the general solution for the generic problem of photography and Eqs. (2.30) and (2.31) provide us with the general solution for the generic problem of task-based imaging.

Note that in case of photography the MTF is almost symmetric around the plane of best focus. As the object reaches either end of the depth of field of interest, the MTF reaches its equal minimum at either of these two ends. In fact our optimization maximizes this minimum by choosing the right cubic phase element. For instance, in the case of the sample problem of Section 2.3, the worst-case $MTF^e$ is 0.10.

However in task-based imaging the MTF is neither symmetric around the best focus, nor do we have the highest MTF at the original best focus. In fact in this case refocusing has removed the symmetry so that we have equal $MTF^e$ at both ends of the desired depth of field *at the maximum spatial frequency of interest.* Obviously this maximum spatial frequency of interest decreases as the object moves toward the imaging system, and thus the MTF at the near field does not need to be

as high as the MTF of the far field. As previously explained, the amount of usable information transfered from the object to the detective device is equal at both ends. Our optimization has maximized this minimum usable information which is transfered at both ends of the depth of field of interest. For instance in the case of the sample problem of Section 2.4, the worst case $MTF^e$ is 0.11.

Note the difference between the problems solved in Sections 2.3 and 2.4. If one wants to use the sample uniform image quality imaging system in Section 2.3 for the sample task-based job in Section 2.4, the worst case MTF would be reduced from 0.11 to 0.10. If one does the reverse, then the worst case MTF is reduced from 0.10 to 0.06. This shows how the system in each case is optimized to do the particular job of interest.

When the design constraints exceed the diffraction limit, the expression for the optimal cubic-phase coefficient becomes a complex value. This can be used as a test of the feasibility of a particular optical design.

Another interesting result which is revealed by this optimization concerns the worst-case MTF in the task-based imaging problem. Although it is expected that moving the image plane changes the worst-case MTF, Eq. (2.30) states that changing the image plane has no effect on the worst-case MTF. The wrong intuition regarding the change of worst-case MTF is because of the magnification change resulting from moving the image plane. Indeed, one might expect that an increase in magnification would call for larger ranges of spatial frequencies of interest for the MTF, thus decreasing the worst-case MTF. However, as it can be seen in Eq. (2.30) one can change the value of the focal length to compensate for that. Intuitively, this fact can be seen as a result of the change of the diffraction limited spatial frequency of the system due to the change of the image plane.

Using our optimal results, the process of pupil function engineering is facilitated enormously. This is not only because now one can get the optimal solution instantly rather than doing a lengthy numerical optimization, but also because now there is an analytic proof that we have reached the approximate global optimum. Note that, although the optimal results presented in this Chapter are based on approximations, we can easily bound the results as near-optimal solution based on our

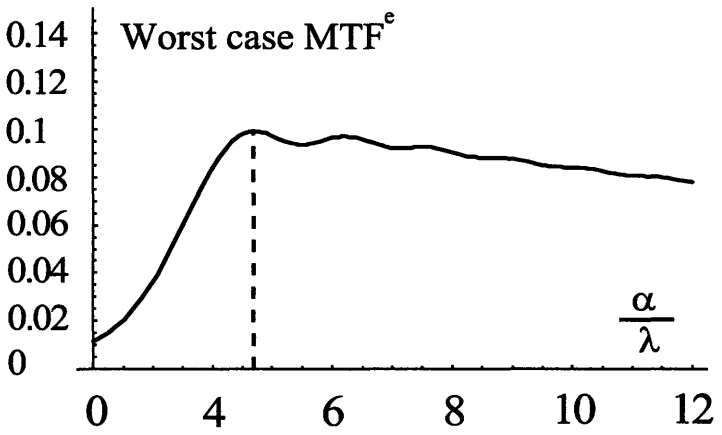



Figure 2-16: Graphical representation of the optimum cubic coefficient (uniform quality imaging problem). This figure is plotted using the optical system specifications provided Table 2.1. It shows how numerical optimization is in accordance with our analytical optimization. The optimum $\alpha$ from the figure is $4.65\lambda$ whereas analytical optimization has shown $\alpha^* = 4.60\lambda$. This difference is the result of the approximations performed in Appendices B, C and D.

accuracy analysis in Appendix B. Roughly speaking, we are not too far from exact global optimum. For instance Fig. 2-16 shows the graph of the $MTF^e(\alpha)$ for the sample problem of Section 2.3. It could be read from this graph that the exact optimum value of $\alpha$ is $4.65\lambda$ as opposed to $4.60\lambda$ that we got in Section 2.3. However as it can be seen in this graph, the actual optimum value of the $MTF^e$ is almost constant for some range of values of $\alpha$. This means that within a small interval close to the approximation value, a slight error in the value of $\alpha$ has little effect on the system performance. As another example, by carefully looking at the solid-line graph in Fig. 2-13 one can see that although Eq. (2.22) holds for approximate MTF ($MTF^{a2}$), it is not satisfied for the exact MTF ($MTF^e$). A slight change in the value of optimum $W_{20}^{'}$ can solve this problem. Thus, these final trims of the optimum value are recommended before using our results for actual design purposes.

In the case of the problem presented in Section 2.3, a good policy is to inspect the values of $\alpha$ close to the one given by Eq. (2.14) in order to find the exact optimum. In the case of task-based imaging, one should first inspect the value of $W_{20}^{'}$ given by Eq. (2.23) [or $d_i$ through Eqs. (2.21) and (2.23)] and find the corresponding exact optimum for which Eq. (2.22) holds. The next step in task-based imaging is to find the exact optimum value of $\alpha$ by inspecting the values close to that of Eq. (2.30).

Another important result in this Chapter is the presentation of an analytical approximated expression for the MTF of an imaging system. Needless to say, having an analytical expression for the MTF speeds up its calculation regardless of the purpose of the calculation. Possible future uses of this expression include but are not limited to image processing, optimization of more complicated merit functions and analysis of the ambiguity function of defocused imaging systems.

# Chapter 3

# Point Spread Function

In this Chapter we introduce a new method for analyzing the diffraction integral for evaluating the point spread function. The new method is based on the use of *higher order Airy functions* along with Zernike and Taylor expansions. Our approach is applicable when we are considering a finite, arbitrary number of aberrations and arbitrarily large defocus simultaneously. We present an upper bound for the complexity and the convergence rate of this method. We also compare the cost and accuracy of this method to traditional ones and show the efficiency of our method through these comparisons. In particular, we rigorously show that this method is constructed in a way that the complexity of the analysis (i.e the number of terms needed for expressing the light disturbance) does not increase as either of defocus or resolution of interest increases. This has applications in several fields that use pupil function engineering such as biological microscopy, lithography and multi-domain optimization in optical systems.

## 3.1 Introduction

The importance of studying the effects of aberrations and defocus on the basis of diffraction theory is very well understood [31] and recent new applications of it, such as biological microscopy [32], lithography [33] and multi-domain optimization techniques in optical systems [18, 20], which need high resolution and accurate value of the point spread function, have called for a more comprehensive

study. For instance, recent articles have reported the use of intentionally added aberrations for making more sophisticated optical systems [21, 18]. Further steps in this direction require a more involved analysis of the diffraction integral in the presence of aberrations and defocus, in order to simplify the process of evaluating the point spread function.

Solving the diffraction integral to find an analytical form for the field distribution on the image plane depends crucially on the defocus and aberration factors. The original Nijboer-Zernike approach for this purpose can only lead to a reasonable approximation when the wavefront deviation due to aberrations and defocus remains within a few radians [31]. Also, even when aberrations and defocus factors are small, but many of them coexist, the Nijboer-Zernike method becomes substantially more complex [31].

Recently, extensions of the original Nijboer-Zernike method have been developed in order to make it applicable to larger values of defocus and aberrations. These expansions lead to a representation of the point spread function whose complexity (i.e. number of terms needed for expressing the point spread function) increases at least linearly with defocus [4, 5, 6, 7].

We present a new method for attacking the diffraction integral problem. Our method provides an expansion for the point spread function with reduced complexity. In particular, the number of terms for expressing the point spread function is uniformly bounded on defocus. This result is demonstrated through rigorous mathematical bounds on the accuracy of the calculated point spread function. Our main result is the following expansion for the point spread function $h$:

$$h(x,y;x_0,y_0) = \sum_{n,m} A'_{nm} \frac{J_{n+1}(R)}{R} \cos[m(\Theta + \phi_0)],$$

where $J_{n+1}(R)$ is the $(n+1)^{\text{th}}$ order first kind Bessel function, $(x,y)$ and $(x_0,y_0)$ are Cartesian coordinate systems at the image and object planes respectively, $R\angle\Theta$ is a polar coordinate system related to those two coordinate systems and $r_0\angle\phi_0$ is the polar coordinate system in the object plane. The coefficients $A'_{nm}$ are polynomials of the aberration constants and of the defocus coefficient multiplied by a factor that is exponential on the defocus coefficient. Functions $\frac{J_n(R)}{R}$ used in the expansion are denoted *higher order Airy functions.* Our method for developing the above repre-

sentation for $h$ is novel and requires a sequence of Taylor and Zernike expansions. The expansions are combined so as to capture the physics of diffraction with a circular aperture. Furthermore, by using the Schwarzschild's representation of the wavefront error, we facilitate the process of investigating the effect of change of aberrations (e.g. primary aberrations) on the defocused point spread function.

In other point spread function expansions in the literature, usually the undefocused wavefront error is represented using Zernike basis functions. This makes evaluation of the point spread function with explicit values for a set of aberrations particularly straightforward [4, 5, 6, 7]. On the other hand, investigating the effect of change of aberrations (e.g. primary aberrations) on the defocused point spread function using these methods is more complicated [5]. This is because one needs to first expand the aberration of interest (e.g. primary aberrations) using Zernike basis functions.

Our expansion for the point spread function exhibits several desirable properties. It can be used to evaluate the point spread function for systems with an arbitrary number of aberrations. It is also computationally tractable and numerically stable over all ranges of defocus values. By taking advantage of the closed-form solution, the diffraction integral may be evaluated within any arbitrary resolution using our expansion. We show that, even though exact representation of $h$ involves an infinite summation of polynomials $A'_{nm}$ of infinite degree, the number of terms and polynomial degree required to achieve a prescribed accuracy scale gracefully with the system parameters. Specifically, we establish an explicit bound showing that, in order to achieve an accuracy of $\epsilon$, the required number of terms grows linearly with the values of aberrations except for defocus, the maximum value of $R$ of interest and $\log \frac{1}{\epsilon}$, and is independent of the remaining parameters of the system. This means that unlike previous methods [4, 5, 6], the complexity of our expansion does not increase as defocus increases. Furthermore, numerical experiments confirm the analytical results obtained.

In the next Section we formally state the problem; this includes the basic assumptions for deriving the diffraction integral and the general aberration form. In Section 3.3, we present the main result which is the general form for the point spread function. There, we consider the most general representation for aberration functions and defocus. In Section 3.4, we analyze the general result

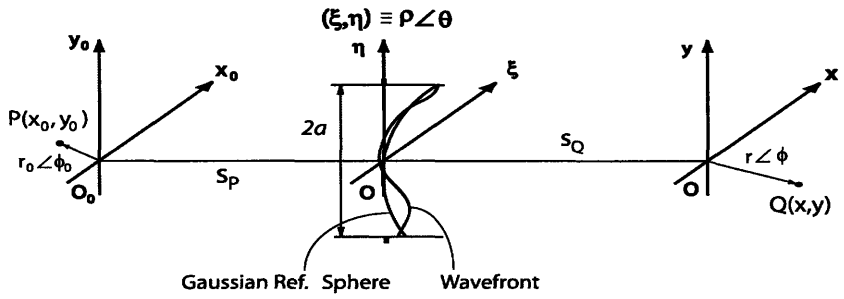



Figure 3-1: Schematic view of the optical system under consideration.

when all primary aberrations and defocus are simultaneously present. In this Section, we also present some examples of point spread function in the case of primary aberrations. In Section 3.5, we analyze the complexity of our method. We present an upper bound for the number of terms and degree of polynomials required in the expansion of $h$ in order to achieve a prescribed accuracy. In Section 3.6, we compare the cost and accuracy of this method to traditional ones and show the efficiency of our method through these comparison. In Appendices F and G, we present a detailed derivation of our method. In Appendix H we present the complexity proofs.

## 3.2 The Optical Point Spread Function

In this Section, we introduce the point spread function (PSF). Figure 3-1 shows the configuration of an arbitrary optical system in the object plane, image plane and pupil plane for computing the PSF. We assume that the usual Sommerfeld-Kirchhoff assumptions hold, and that the chromatic aberrations are negligible. The PSF $h$ is used to calculate the image disturbance $\vec{Q}$ caused by a monochromatic coherent plane wave illumination in an arbitrary plane parallel to the exit pupil in the presence of an object $\vec{P}$. In particular, at each point $(x, y)$ on the image plane, we have

$$\vec{Q}(x,y) = \int_{\mathcal{A}} h(x,y;x_0,y_0)\,\vec{P}(x_0,y_0)\,dx_0\,dy_0, \tag{3.1}$$

where $\mathcal{A}$ is the whole object domain in the object plane.

The PSF $h$ can be further specified as follows: Consider a point source of monochromatic light

$P$ and find the disturbance in an arbitrary point $Q$ in space, assuming a circular aperture of radius $a$. Let $(x_0, y_0)$ denote the ray entrance Cartesian coordinates on the object plane at distance $S_P$ from the entrance pupil and let $r_0 \angle \phi_0$ represent the respective polar coordinates. According to Huygens-Fresnel principle [31], the disturbance at an arbitrary point $(x, y)$ (or in polar coordinates $r \angle \phi$) on the image plane at distance $S_Q$ from the exit pupil is

$$h(x, y; x_0, y_0) = C \int_0^{2\pi} \int_0^1 \mathrm{e}^{\mathrm{i}\, k\, w(\rho, \theta, r_0, \phi_0)} \mathrm{e}^{\mathrm{i} R \rho \cos(\theta - \Theta)} \rho \, d\rho \, d\theta. \tag{3.2}$$

The image plane is not necessarily the Gaussian image plane, which is at distance $S_G$ of the lens. In this formulation, $\rho$ and $\theta$, which are integration variables, are polar coordinates in the exit pupil plane. Coordinates $R$ and $\Theta$ are polar equivalents of the point $(u, v)$, which is related to $(x_0, y_0)$ and $(x, y)$ according to

$$u = -k\, a \left( \frac{x_0}{r'} + \frac{x}{s'} \right), \tag{3.3}$$

$$v = -k\, a \left( \frac{y_0}{r'} + \frac{y}{s'} \right), \tag{3.4}$$

$$r'^2 = {x_0}^2 + {y_0}^2 + {S_P}^2, \tag{3.5}$$

$$s'^2 = x^2 + y^2 + {S_Q}^2, \tag{3.6}$$

where $k = 2\pi/\lambda$ is the wave number. The wavefront error, $w$, is the deviation of the wavefront from the Gaussian reference sphere in the exit pupil. It includes all aberrations and defocus terms.

It can be shown that the coefficient $C$ in Eq. (3.2) is [31]

$$C = \frac{\mathrm{i}\, k \cos(\delta)}{2\pi r' s'}. \tag{3.7}$$

where $\delta$ is defined as the acute angle which satisfies

$$\tan(\delta) = \frac{\sqrt{(x_0 + x)^2 + (y_0 + y)^2}}{S_P + S_Q}. \tag{3.8}$$

Note that $C$ is bounded in the whole region of integration as

$$|C| \leq \frac{k}{2\pi |S_P||S_Q|}. \tag{3.9}$$

Thus, to attack the main problem of finding an analytic solution to the diffraction integral, we may neglect the coefficient $C$ in Eq. (2), and define $\hat{h}$, the normalized PSF, as

$$\hat{h}(x, y; x_0, y_0) = \frac{1}{2\pi} \int_0^{2\pi} \int_0^1 e^{i\,k\,w(\rho,\theta,r_0,\phi_0)} e^{iR\rho\cos(\theta-\Theta)} \rho \, d\rho \, d\theta. \tag{3.10}$$

### 3.2.1 Schwarzschild's Aberration Coefficients

In general, in terms of optical path length, $w$ is a function of the source coordinates $(r_0 \angle \phi_0)$ and the pupil coordinates $(\rho \angle \theta)$. The particular dependence of $w$ on these four variables depends on the properties of the optical systems under consideration. For a rotational symmetric optical system, it is easy to show that $w$ is only a function of $\theta - \phi_0$ rather than $\theta$ and $\phi_0$ independently. Furthermore, since the analysis of the point spread function is done for a particular object point (i.e. the integrals in Eq. (3.2) are over $\rho$ and $\theta$), we are not interested in the particular dependence of $w$ on $r_0$ at this point. Thus, we are left with only two sets of variables, namely $\rho$ and $\theta - \phi_0$. Now considering the fact that $w$ is in fact the deviation of the wavefront from the Gaussian reference sphere, using perturbation theory, Schwarzschild has shown that we can express $w$ as [34]

$$w(\rho, \theta, r_0, \phi_0) = \sum_{l,m=0}^{\infty} f_{l,m}(r_0) \rho^{2l+m} \cos^m(\theta - \phi_0). \tag{3.11}$$

Note that when the wavefront is close to the Gaussian reference sphere this representation of the wavefront error requires the minimum number of terms for a prescribed accuracy for representing $w$. Also note that in Eq. (3.11), we have only shown the dependence of $w$ on $r_0$ in the functional form $f_{l,m}(r_0)$. As explained before, this is because we are not interested in the particular form of $f_{l,m}$. The functions $f_{l,m}$ are referred to as the aberration coefficients. The particular form of $f_{l,m}$ depends on the optical system configuration and properties. The dimension of the aberration coefficient $f_{l,m}$ is $L^{2l+m-1}$, where $L$ refers to length dimension. Note that this dimension for aberration coefficients

ensures that the dimension for wavefront error $w$ is length as expected from its definition. Note that in practice it is often sufficient to consider the first five terms for aberrations, referred to as primary aberrations ($(l,m) \in \{(0,1),(0,2),(1,0),(1,1),(2,0)\}$).

Following Schwarzschild's analysis [34], we have

$$w(\rho,\theta,r_0,\phi_0) = \sum_{j=1}^{n_{ab}} \left\{ f_{L_j,M_j}(r_0)\,(a\,\rho)^{2L_j}\,\left[a\,\rho\,\cos(\theta-\phi_0)\right]^{M_j} \right\}, \tag{3.12}$$

In Eq. (3.12), $n_{ab}$ is the total number of aberrations under consideration. Note that the particular value of $L_j$ and $M_j$ identifies the type of aberration which $j$ is referring to. In particular $f_{1,0}$ or $DF$ is referred to as the defocus coefficient. We treat defocus separately in order to make the complexity of the expansion independent of defocus. The particular functional form of defocus is given by:

$$f_{L_1,M_1}(r_0) = f_{1,0}(r_0) = DF = \frac{1}{2}\left(\frac{1}{S_Q} - \frac{1}{S_G}\right). \tag{3.13}$$

Comparing Eqs. (3.11) and (3.12) one notes that we have replaced $\rho$ with $a\rho$ in Eq. (3.12), where $a$ is the aperture radius. This is done to keep $\rho$, the exit pupil's radial coordinate, in normalized form.

If we seek the point spread function of a known optical system, the numerical evaluation of the aberration coefficients becomes important. The numerical value of the aberration coefficients are usually found using ray tracing packages (like Zemax); however for simple optical systems, geometrical analysis could be performed to find the exact value of the aberration coefficients. In general if have a set of data points for $w$, using a standard curve fitting method we can find the aberration coefficients. Having a set accuracy and limited data points, we only consider a finite number of terms in the expression of $w$.

An important application of having an expression for PSF is in pupil function engineering design [21, 18, 20, 13]. There, since the goal is to find the appropriate set of intentionally induced aberrations, there is no need to evaluate the aberration coefficients using a set of data points. In fact in that case, aberration coefficients are determined through the pupil function engineering process

using the proposed expansion for PSF. We find the total wavefront error using Eq. (3.12). Then we fabricate our pupil function engineering element such that it induces the corresponding wavefront error in the system.

In this Chapter we develop an expansion for $\hat{h}$ in terms of polynomials, in the presence of aberrations and defocus. The expansion involves an infinite sum of polynomials, but we show that, for any given accuracy, only a finite number of terms is required.

## 3.3 The PSF Expansion for Arbitrary Wavefront Errors and Defocus

We now present a general expression for the PSF as an expansion in terms of *higher order Airy functions*. We define the $n^{\text{th}}$ order Airy function as

$$n^{th} \text{ order Airy function} = \frac{J_{n+1}(R)}{R},$$

where $J_{n+1}$ is the $(n+1)^{\text{th}}$ order first kind Bessel function.

We represent the PSF as a sum of polynomials of the aberration and defocus coefficients. In this Section, finitely many of the aberration terms in Schwarzschild's analysis are considered. In practice, however, only a few of those (usually the primary aberrations) are of real importance. We illustrate application of our result in one such case in the next Section.

Our proposed expansion is of the form

$$\hat{h}(x,y;x_0,y_0) = \sum_{n=0}^{\infty}\sum_{m=0}^{n}\left\{\tilde{\delta}_{n-m}\delta_m A_{nm}\cos\left[m(\Theta+\phi_0)\right]\frac{J_{n+1}(R)}{R}\right\}, \tag{3.14}$$

where $\tilde{\delta}_i$, $\delta_i$ and the coefficients $A_{nm}$ are given by Eqs. (3.15), (3.16) and (3.17) respectively.

$$\tilde{\delta}_i \;=\; \begin{cases} 1 & \text{if } i \text{ is even,} \\ 0 & \text{otherwise.} \end{cases} \tag{3.15}$$

$$\delta_i \;=\; \begin{cases} 1 & \text{if } i = 0, \\ 2 & \text{otherwise.} \end{cases} \tag{3.16}$$

$$A_{nm} \;=\; \frac{n+1}{2^{m-1}}\, \mathrm{e}^{\mathrm{i}\, m\phi_0} \sum_{(\mathbf{N},D)\in\aleph_m} D\, \beta_{\mathbf{N}}\, S^m_{n,k_{\mathbf{N}}}(\vec{\beta}). \tag{3.17}$$

$\aleph_m$ is a set of pairs of vectors $\mathbf{N} = [N_2,\, N_3,\, \ldots\, N_{n_{ab}}]$ and scalars $D$ defined as

$$\aleph_m = \left\{ (\mathbf{N}, D) \;|\; \sum_{j=2}^{n_{ab}} (M_j N_j) = m + 2k,\; D = \frac{(m+2k)!}{2^{2k} k! (m+k)!};\; k, N_j \in \mathcal{N} \right\}. \tag{3.18}$$

where $M_j$ is the proper parameter used in the definition of the $j^{th}$ aberration as in Eq. (3.12) and $\mathcal{N}$ is the set of natural numbers. Function $S^m_{n,k_{\mathbf{N}}}(\vec{\beta})$ is given by

$$S^m_{n,k_{\mathbf{N}}}(\vec{\beta}) = \int_0^1 \prod_{j\in\chi_1} \mathrm{e}^{\beta_j \rho^{2L_j}}\, R^m_n(\rho)\, \rho^{k_{\mathbf{N}}+1}\, d\rho, \tag{3.19}$$

where $R^m_n(\rho)$ is the Zernike polynomial introduced in Appendix F, and we have

$$\beta_j = \mathrm{i}\, k\, f_{L_j,M_j}(r_0)\, a^{2L_j+M_j}, \tag{3.20}$$

$$\beta_{\mathbf{N}} = \prod_{j\in\chi_2} \frac{(\beta_j)^{N_j}}{N_j!}, \tag{3.21}$$

$$k_{\mathbf{N}} = \sum_{j\in\chi_2} (2L_j + M_j) N_j, \tag{3.22}$$

$$\chi_1 = \{j \,|\, M_j = 0,\, j = 1, \ldots, n_{ab}\}\,, \tag{3.23}$$

$$\chi_2 = \{j \,|\, M_j \neq 0,\, j = 1, \ldots, n_{ab}\}\,, \tag{3.24}$$

$$\chi_3 = \{j \,|\, M_j = 0,\, j = 2, \ldots, n_{ab}\}\,. \tag{3.25}$$

Note that $\beta_j$ can be considered as the final aberration coefficient. It is a dimensionless variable. This removes any confusion in expressing its value. In particular, $\beta_1$, the final defocus coefficient, is also a dimensionless quantity. In some literature $a^2 f_{1,0} = a^2 DF$ is referred to as the defocus coefficient. Note that $a^2 DF$ has a dimension of length and may be expressed in *micrometers*, number of wavelengths or simply *meters*.

A derivation of the expressions above can be found in Appendix F. Note that $S^m_{n,k_{\mathbf{N}}}(\vec{\beta})$ is defined implicitly in Eq. (3.19), requiring computation of an integral. An explicit expression for the integral, which is based on a Taylor expansion of $e^{\beta_j \rho^{2L_j}}$ (for $j \in \chi_3$), can be found in Appendix G. The derivation is tedious but relatively straightforward. It follows from this expansion that $S^m_{n,k_{\mathbf{N}}}(\vec{\beta})$ can be expressed as a polynomial of aberration constants $\vec{\beta}$ multiplied by a factor that is exponential on the defocus coefficient.

The number of terms in the summation in Eq. (3.14) and the degree of the polynomials used to express $S^m_{n,k_{\mathbf{N}}}(\vec{\beta})$ are infinite. However, in Section 3.5 we show that, for any desired accuracy $\epsilon$, a finite truncation of Eq. (3.14) as well as finite-degree polynomials for $A_{nm}$ in Eq. (3.17) suffice for an appropriate approximation to $\hat{h}$; i.e. Eq. (3.14) converges to Eq. (3.10). We give an explicit bound on the number of terms and degree required and we show that they scale gracefully with the systems parameters and $\epsilon$. This will be realized by giving an upper bound for $n$ in Eq. (3.14) as well as an upper bound for every $N_j$ in Eq. (3.18). Note that a bound on $N_j$ will determine the number of terms of Taylor expansion of $e^{\beta_j \rho^{2L_j + M_j} [\cos(\theta - \phi_0)]^{M_j}}$ which have been used in our expansion.

In the next Section we illustrate some of the applications of this expansion through examples.

## 3.4 Examples

In this Section we consider the primary (Seidel) aberrations and defocus ($n_{ab} = 5$):

$$\mathbf{i}\,k\,a^{2L+M}\,f_{L,M}(r_0) = \begin{cases} \gamma_1 & \text{if (L,M)=(1,0) } \textit{Defocus and Field Curvature,} \\ \gamma_2 & \text{if (L,M)=(2,0) } \textit{Spherical Aberration,} \\ \gamma_3 & \text{if (L,M)=(0,1) } \textit{Distortion,} \\ \gamma_4 & \text{if (L,M)=(0,2) } \textit{Astigmatism,} \\ \gamma_5 & \text{if (L,M)=(1,1) } \textit{Coma,} \end{cases} \tag{3.26}$$

where for simplicity $(x_0, y_0)$ is assumed to be $(0,0)$. Note that all the final primary aberration coefficients, $\gamma_1 \ldots \gamma_5$, are dimensionless. Substituting in Eq. (3.17) we have

$$\begin{aligned} A_{nm} &= \frac{n+1}{2^{m-1}}\,\mathbf{i}^n \times \\ &\sum_{(\mathbf{N},D)\in\aleph_m} D\left[\frac{\gamma_3^{N_3}\gamma_4^{N_4}\gamma_5^{N_5}}{N_3!N_4!N_5!}\,S^m_{n,N_3+2N_4+3N_5}(\gamma_1,\gamma_2)\right], \end{aligned} \tag{3.27}$$

where

$$\begin{aligned} \aleph_m &= \{(\mathbf{N},D) = (N_3,N_4,N_5,D)\,| \\ &N_3 + 2N_4 + N_5 = m + 2k,\; D = \frac{(m+2k)!}{2^{2k}k!(m+k)!};\; k, N_3, N_4, N_5 \in \mathcal{N}\Big\}, \end{aligned} \tag{3.28}$$

$$S^m_{n,N_3+2N_4+3N_5}(\gamma_1,\gamma_2) = \int_0^1 \mathrm{e}^{\gamma_1\rho^2+\gamma_2\rho^4}\,R^m_n(\rho)\,\rho^{N_3+2N_4+3N_5+1}\,d\rho, \tag{3.29}$$

and the derivation of $S$ can be found in Appendix G. Thus $A_{nm}$ is a polynomial of $\gamma_1, \ldots, \gamma_5$. It also has one term in the form of $\exp(\gamma_1)$. Although from the above equation it seems that the order of this polynomial is infinity, as will be explained later, once we set a target accuracy, all except for a few terms in $A_{nm}$ become negligible. As it will be shown in Section 3.5, the number of necessary terms in expression (3.27) scales favorably with the desired accuracy of the representation.

Now to evaluate the transfer function, $\hat{h}$, we can rewrite Eq. (3.14) as

$$\hat{h}(r\angle\phi) = \sum_{n=0}^{\infty}\sum_{m=0}^{n}\left\{\tilde{\delta}_{n-m}\delta_m A_{nm}\cos[m(\phi+\pi)]\frac{J_{n+1}(Br)}{Br}\right\}, \tag{3.30}$$

where $B = \frac{k\alpha}{s'}$ is obtained using Eqs. (3.3) and (3.4) and by setting $(x_0, y_0)$ equal to $(0,0)$. As an example, the results using this method are shown in Fig. 3-2 for $Br < 20$. $\gamma_2$ and $\gamma_3$ are zero in this figure. Thus, we have (note that since $\gamma_3 = 0$, $(\mathbf{N}, D)$ is a three element vector)

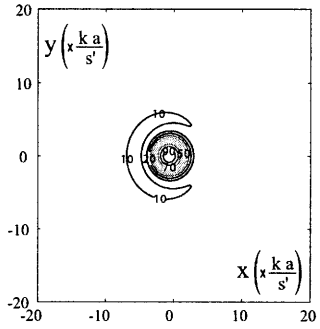
$$\begin{aligned}
\aleph_0 &= \Big\{(0,0,1),\ (2,0,\tfrac{1}{2}),\ (0,1,\tfrac{1}{2}),\ (4,0,\tfrac{3}{8}),\ (2,1,\tfrac{3}{8}),\ (0,2,\tfrac{3}{8}), \\
&\quad (4,1,\tfrac{5}{16}),\ (2,2,\tfrac{5}{16}),\ (0,3,\tfrac{5}{16}),\ (4,2,\tfrac{35}{128}),\ (2,3,\tfrac{35}{128}), \\
&\quad (0,4,\tfrac{35}{128}),\ (4,3,\tfrac{63}{256}),\ (2,4,\tfrac{63}{256}),\ (4,4,\tfrac{231}{1024})\Big\}, \\
&\quad \vdots \\
\aleph_{11} &= \{(3,4,1)\}, \\
\aleph_{12} &= \{(4,4,1)\},
\end{aligned} \tag{3.31}$$

and $\aleph_m = \{\}$ otherwise (where we have assumed $N_j \le N_j^* = 4$). This means that $A_{nm}$ is zero for $m > 12$. Also in this special case, Eq. (3.29) reduces to

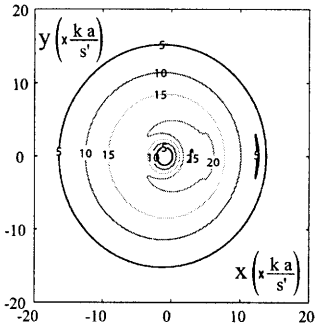
$$S^m_{n,k'}(\gamma_1) = \sum_{l=0}^{(n-m)/2}\left(\frac{C^m_{n,l}}{2}(-\gamma_1)^{\frac{-(2+n-2l+k')}{2}}(\frac{n-2l+k'}{2})!\left[1-e^{\gamma_1}\sum_{j=0}^{\frac{n-2l+k'}{2}}\frac{(-\gamma_1)^j}{j!}\right]\right), \tag{3.32}$$

where $C^m_{n,l}$ is defined

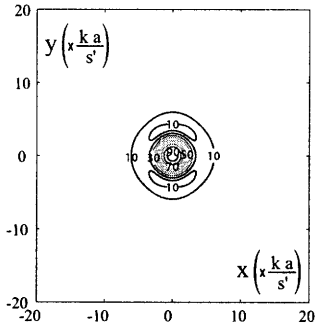
$$C^m_{n,l} = \frac{(-1)^l(n-l)!}{l![(n+m)/2-l]![(n-m)/2-l]!}. \tag{3.33}$$



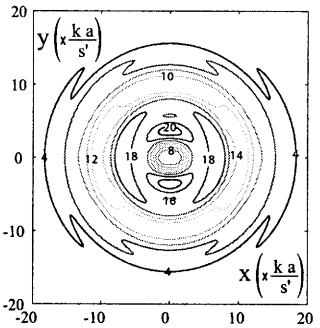

(a) $\gamma_1 = 0$, $\gamma_4 = 0$, $\gamma_5 = 1\mathbf{i}$.



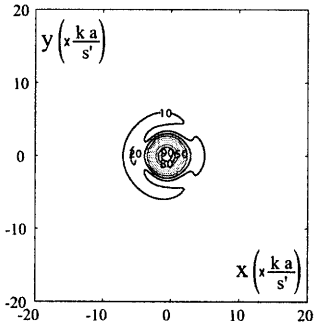

(b) $\gamma_1 = 2\pi\mathbf{i}$, $\gamma_4 = 0$, $\gamma_5 = 1\mathbf{i}$.



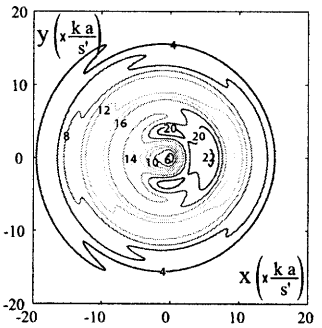

(c) $\gamma_1 = 0$, $\gamma_4 = 1\mathbf{i}$, $\gamma_5 = 0$.




(d) $\gamma_1 = 2\pi\mathbf{i}$, $\gamma_4 = 1\mathbf{i}$, $\gamma_5 = 0$.




(e) $\gamma_1 = 0$, $\gamma_4 = 1\mathbf{i}$, $\gamma_5 = 1\mathbf{i}$.




(f) $\gamma_1 = 2\pi\mathbf{i}$, $\gamma_4 = 1\mathbf{i}$, $\gamma_5 = 1\mathbf{i}$.

Figure 3-2: Contour plot of the modulus of PSF, $|h|$, in the presence of aberrations and defocus (normalized to 100).

## 3.5 Complexity Analysis

In this Section we analyze the complexity of our representation of the PSF. Specifically, we show rigorously that within a confined region of space (i.e. the exit window) the PSF can be expressed within any arbitrary accuracy, using a finite number of terms in Eq. (3.14) regardless of the value of defocus and desired resolution (note that by resolution, we mean the shortest distance between two points where we are interested to evaluate PSF). This means that, as we increase the resolution of interest or as we change the defocus, the number of necessary terms within the prescribed accuracy do not change. This is of great importance in many practical cases where numerical simulation fails to generate the PSF within the required resolution and accuracy in a reasonable time. This issue is revisited in the next Section.

Considering a desired accuracy, the complexity of the expansion in Eq. (3.14) depends on three factors: (i) The maximum index of summation, $n^*$, considered in Eq. (3.14). (ii) The number, $N^*$, of terms in the summation considered in Eq. (3.14); this number is $O\left((n^*)^2\right)$. (iii) The degree of polynomials involved in the expressions of $A_{nm}$ in Eq. (3.17). These polynomials are at most on the order of ${N_j}^*$ on $\beta_j$, the $j^{th}$ aberration coefficient, when $N_j \leq {N_j}^*$ in Eq. (3.18). We analyze all these three factors.

With the finite summation bound $n^*$, and the finite polynomial order ${N_j}^*$ for each aberration coefficient $\beta_j$, $j = 2, \ldots, n_{ab}$, Eq. (3.14) is rewritten as

$$\hat{h}_{n^*}(x, y; x_0, y_0) = \sum_{n=0}^{n^*} \sum_{m=0}^{n} \tilde{\delta}_{n-m} \delta_m A^*_{nm} \cos[(\Theta + \phi_0)] \frac{J_{n+1}(R)}{R}, \tag{3.34}$$

where $A^*_{nm}$ is defined as

$$A^*_{nm} = \frac{n+1}{2^{m-1}} e^{im\phi_0} \sum_{(\mathbf{N}, D) \in \aleph^*_m} D\, \beta_{\mathbf{N}}\, S^m_{n, k_{\mathbf{N}}}(\vec{\beta})^*, \tag{3.35}$$

and $\aleph^*_m$ and $S^m_{n,k_{\mathbf{N}}}(\vec{\beta})^*$ are defined as

$$\aleph_m^* = \left\{ (\mathbf{N}, D) \mid \sum_{j=2}^{n_{ab}} (M_j N_j) = m + 2k,\ D = \frac{(m+2k)!}{2^{2k} k! (m+k)!};\ N_j \leq N_j^*,\ k, N_j \in \mathcal{N} \right\}, \tag{3.36}$$

$$S_{n,k_{\mathbf{N}}}^m(\vec{\beta})^* = \sum_{l=0}^{(n-m)/2} C_{n,l}^m \int_0^1 \mathrm{e}^{\beta_1 \rho^2} \prod_{j \in \chi_3} \left[ \sum_{N_j=0}^{N_j^*} \frac{(\beta_j \rho^{2L_j})^{N_j}}{N_j!} \right] \rho^{n-2l+k_{\mathbf{N}}+1}\, d\rho. \tag{3.37}$$

Note that the only difference in the definition of $\aleph_m^*$ and $S_{n,k_{\mathbf{N}}}^m(\vec{\beta})^*$ and $\aleph_m$ and $S_{n,k_{\mathbf{N}}}^m(\vec{\beta})$ is that $N_j$ is bounded by $N_j^*$ in $\aleph_m^*$ and $S_{n,k_{\mathbf{N}}}^m(\vec{\beta})^*$.

As the accuracy of interest in Eq. (3.34) increases, the upper bounds for $n$ and $N_j$, i.e. $n^*$ and $N_j^*$, should also increase. The change of these bounds as the desired accuracy in Eq. (3.34) changes is an expression of complexity of our expansion. Theorem 3.5.1 provides us with such an expression, and is our main result in this Section.

**Theorem 3.5.1.** *Let $\epsilon$, $n_{ab}$ and $R^*$ be arbitrary and let*

$$n^* \geq \max\left(5,\ \mathrm{e}R^* + 1,\ 2\log_2 \frac{1}{\mathrm{e}(2\mathrm{e}-1)\sqrt{\pi}\epsilon}\right),$$

*and*

$$N_j{}^* \geq \max\left(4,\ 2\mathrm{e}\,|\beta_j| + 1,\ \log_2 \frac{\sqrt{6\mathrm{e}}\,\mathrm{e}^3\, n_{ab}\,(1 + R^{*4/3})}{\pi(2\mathrm{e}-1)\epsilon}\right),$$

*for all $j = 2, \ldots, n_{ab}$. Then we have*

$$|\hat{h}(x, y; x_0, y_0) - \hat{h}_{n^*}(x, y; x_0, y_0)| \leq \epsilon$$

*for all $x$, $y$, $x_0$, $y_0$ such that the corresponding value of $R$ is less than or equal to $R^*$.*

Theorem 3.5.1 provides an upper bound on the minimum necessary number of terms in Eqs. (3.34) and (3.35) and proves that it is finite. In fact, numerical simulations in practice suggest that

on average even fewer terms may suffice. A proof of Theorem 3.5.1 along with an intuitive discussion about it can be found in Appendix H.

We have shown that any arbitrary accuracy of the light disturbance in the circle of $R \leq R^*$ can be achieved with a sufficiently large finite value of $n^*$ and $N_j^*$s. Theorem 3.5.1 states that as the radius of the region of interest, $R^*$, increases, the maximum necessary index of summation in Eq. (3.34), $n^*$, increases linearly with $R^*$. It also indicates that the maximum necessary index of summation in Eq. (3.34), $n^*$, increases proportionally to $\log \frac{1}{\epsilon}$, where $\epsilon$ is the accuracy of approximation. We can also see that the maximum necessary index of summation in Eq. (3.35) (as stated in Eqs. (3.36) and (3.37)), $N_j^*$, or in other words, the maximum order of $\beta_j$ in the expression of coefficients $A_{nm}$, increases linearly with the corresponding aberration coefficient $|\beta_j|$ and $\log \frac{1}{\epsilon}$, where $\epsilon$ is the accuracy of approximation. The $\log \frac{1}{\epsilon}$ dependence of $n^*$ and $N_j^*$ on the accuracy ($\epsilon$) confirms the fast convergence of this method.

Considering the above analysis, we conclude that when we are interested in the disturbance in a confined region, we only need to consider a few terms in Eq. (3.14). Now we can move on to the second factor, i.e. $N^*$. To find the total number of terms necessary for a desired accuracy, we recall that Eq. (3.14) has the structure of Zernike polynomials; i.e. $n \geq m$, $n, m > 0$, and $n - m$ =even. Using elementary number theory, one can conclude that the total number of necessary terms in Eq. (3.14) is

$$N^* = \left\lfloor \frac{n^* + 2}{2} \right\rfloor \left\lceil \frac{n^* + 2}{2} \right\rceil. \tag{3.38}$$

Apparently, the number of terms in $A_{nm}$ depends on $\aleph_m$ and $S^m_{n,k_{\mathbf{N}}}(\beta)$, which both in turn depend on the value of $N_j^*$s. This is due to the Taylor expansion that we have used. Using the analysis in Appendix F and the values of $N_j^*$, we can determine the complexity of the $A_{nm}$. The coefficients $A_{nm}$ are polynomials of the aberration constants of order no more than $N_j^*$ for each particular aberration coefficient. The coefficients $A_{nm}$ also depend on the defocus coefficient both in the form of rational polynomial of order no more than $1 + (n + m)/2 + \sum_{j \in \chi_4} N_j{}^*$ and in the

form of $\exp(\beta_1)$, where $\chi_4$ is given by Eq. (3.39). Hence, it is clear that increasing defocus does not increase the complexity of coefficients $A_{nm}$ in a confined region of interest.

$$\chi_4 = \{j | L_j \neq 0, j = 2, \ldots, n_{ab}\} \tag{3.39}$$

The above analysis gives us a comprehensive understanding of an upper bound on the complexity of calculating the light disturbance within the exit window $R^*$. These bounds are representative of the worst case scenario. Numerical experiments, however, suggest that our method on average works better than what analytical bounds suggest. For instance, for the case of $R^* = 40$ and $\epsilon = 0.001$, using Theorem 3.5.1, $n^* = 81$, whereas experimental result suggests $n^* = 45$. Nevertheless, Theorem 3.5.1 is the tightest theoretical bound currently available.

Performing the same experiment for different values of $R^*$ suggests that $n^* = \lceil R^* \rceil + 5$ suffices for $\epsilon = 0.001$. Replacing $n^*$ in Eq. (3.38) by its experimental value , i.e. $\lceil R^* \rceil + 5$, one can get the following expression for the total number of necessary terms in Eq. (3.14) (or Eq. (3.34)) for an accuracy of $\epsilon = 0.001$ in a desired range $R^*$

$$N^* = \left\lfloor \frac{\lceil R^* \rceil + 7}{2} \right\rfloor \left\lceil \frac{\lceil R^* \rceil + 7}{2} \right\rceil. \tag{3.40}$$

The above two equations show the necessary number of terms to express the diffraction integral within a desired range and accuracy. This is of much greater importance when we recall that the number of terms required in the expansion is independent of the values of aberrations and defocus and the required resolution. In other words, regardless of the properties of the imaging system, the above number of terms is sufficient for calculating the light disturbance in the image plane. For instance for an optical system with $f = 50mm$, $f/\# = 3mm$ and pixel-size= $5\mu m$, if we consider a circle with radius of 5 pixels around each pixel and accuracy of $\epsilon = 0.001$, then $R^*$ is 47.5 and thus we do not need to consider terms with $n > 53$ no matter how large our defocus or aberrations are

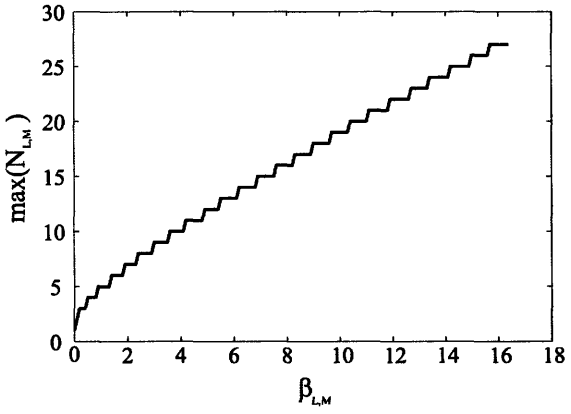



Figure 3-3: Variation of the partial number of terms necessary with $\beta_{L,M}$ for $\epsilon = 0.001$ and $R^* = 20$.

or how fine our resolution is.

We have also performed experiments for finding the minimum number of Taylor expansion terms necessary for each aberration, $N_j^*$, for an accuracy of $\epsilon = 0.001$ and range of interest of $R^* = 20$. These results are shown in Fig. 3-3. One can notice the gap between the theoretical and experimental bounds by comparing Theorem 3.5.1 and Fig. 3-3. For instance for $|\beta| = 5$, using Theorem 3.5.1 one gets $N_j^* = 29$, whereas experimental results suggest $N_j^* = 11$.

Note that without considering the number of aberrations present and their range of values, we cannot state a general result about the absolute or relative errors of Eq. (3.34) in the whole infinite image plane, i.e when $R^* \rightarrow \infty$. For instance, when the distortion aberration coefficient ($\gamma_3$) is large, the PSF peak can shift out of the exit window, causing the absolute and relative approximation errors to increase without bound. This example is shown in Fig. 3-4.

By this complexity analysis we presented bounds on the error of our PSF representation. The presented analytic accuracy expressions along with the performed numerical experiments can provide us with a better vision of the behavior of our PSF representation. In other words while the numerical experiments give us an idea of how well our method works on average, analytic accuracy expressions provide us with a bound on the worst case scenario.

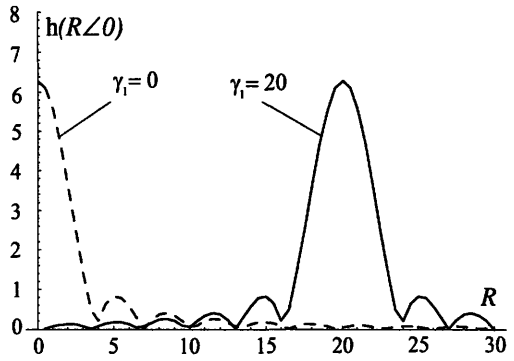



Figure 3-4: Radial variation of the modulus of PSF with and without Distortion (normalized to $2\pi$).

## 3.6 Discussions

Eqs. (3.14) and (3.17) are general expressions for the study of the effect of aberrations and defocus on PSF on the basis of diffraction theory. Two important points that are hidden in these equations are their ability to handle high defocus cases without facing any numerical problems and the potential of this method to consider the effect of as many aberrations as needed at the same time as defocus. In fact any arbitrary aberration can be approximated using Eq. (3.12) and then its effect on the imaging system will be immediately available.

This latter property is very useful in pupil function engineering [21, 18, 20, 13]. In this technique we use general aberrated optical elements (traditionally aspheric) and digital post processing together to increase the performance and/or decrease the cost of imaging systems.

Another important fact about this method is its fast performance compared to direct calculation. It is often the case that direct ray tracing does not suffice for practical needs and one has to analyze the effect of aberrations and defocus using diffraction theory. In that case, our method proves to be very efficient. Figure 3-5 shows the time required to evaluate the PSF at 400 different points in the image plane. It can be seen that for all values of aberrations and defocus, the time required by the new method is significantly less. The ratio of time needed varies from 150 for zero defocus to more than 20000 for defocus of $\beta_1 = 11424\mathrm{i}$ (or $a^2DF = 1000\mu m = 1818.181\lambda$, where $\lambda$ is assumed to be $0.55\mu m$). It should be noted that in all of the calculations in this figure, the accuracy has been kept

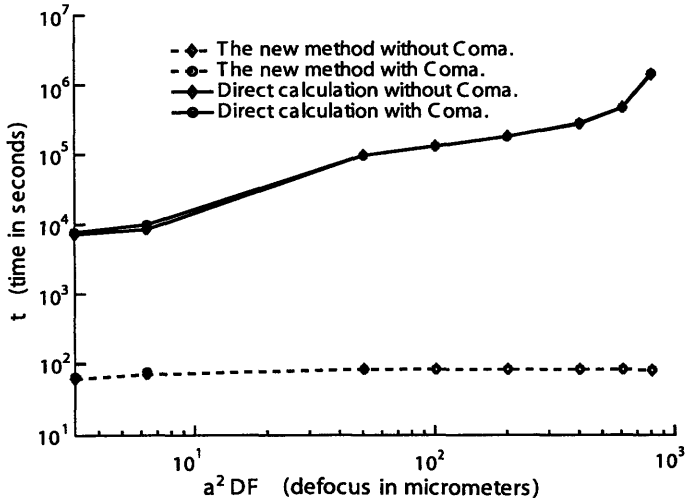



Figure 3-5: Time required for evaluating PSF at 400 different points v.s. defocus ($\epsilon = 0.1\%$).

at 0.1%.

It is to be noted that the traditional method mentioned in Fig. 3-5 is the direct evaluation of the diffraction integral. It is a common practice to use fast Fourier transform (FFT) rather than direct integration, to enhance the speed of calculation. Although the FFT method is almost invariant to defocus and aberration coefficient values, it fails to perform well as the resolution of interest increases. This is illustrated in Figure 3-6. There, the performance of our method and FFT are compared as resolution of interest increases. Note that in this figure the accuracy is 10%; this is because FFT needs too much memory and CPU time for higher accuracies.

The results shown in Figures 3-5 and 3-6 are direct consequences of the complexity properties of our method. In other words, since the complexity of the method does not increase with the increase of defocus or resolution, the time required for calculating the light disturbance does not increase either. The detailed and concrete complexity analysis in Section 3.5 not only provides a theoretical guarantee for the accurate performance of our method, but also presents a reliable tool for evaluating the complexity of a particular PSF calculation task through Theorem 3.5.1.

In fact, Theorem 3.5.1 states that for any arbitrary region of interest and any arbitrary accuracy, there exists a maximum necessary index of summation for Eq. (3.14) and maximum necessary degree of polynomials for Eq. (3.17) which together provide us with an accurate approximation within the

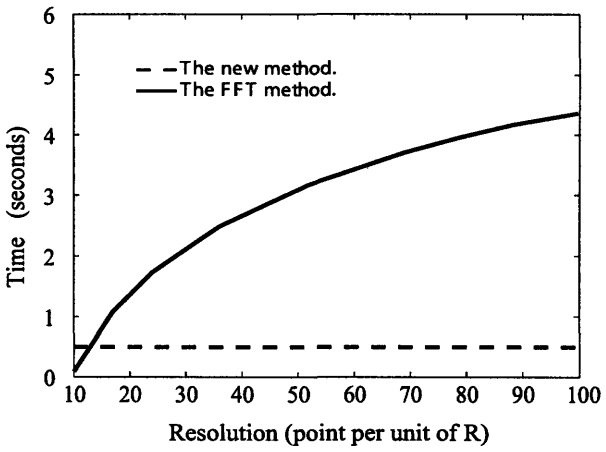



Figure 3-6: Time required for evaluating PSF v.s. resolution ($\epsilon = 10\%$).

region of interest. This theorem also supplies an upper bound for the maximum necessary index of summation and maximum necessary degree of polynomials. Nevertheless numerical experiments suggest that, these bounds are representative of the worst case scenario and our method on average works better than what analytical bounds suggest.

Note that in the derivation of our representation of PSF we have assumed a unit transmittance pupil function. In general one may think of an imaging system with a nonuniform pupil transmittance function, $A(\rho, \theta)$, where $\rho$ and $\theta$ are the pupil coordinates. Thus, the complexity of the final representation of the point spread function would inevitably depend on the complexity of $A$ [4]. Furthermore the complexity of the PSF may no longer be independent of defocus. To investigate the effect of nonuniformity of pupil function on our representation of the point spread function remains a part of our future work. At present, the only approaches that are able to deal with nonuniform pupil functions exhibit complexity that increases with defocus [5, 4, 6, 7]; it remains to be seen whether independence on defocus can be achieved.

Another interesting property of our expansion is its advantage in facilitating the process of calculating the amplitude transfer function (ATF), optical transfer function (OTF), and modulation transfer function(MTF) for an imaging system. This is due to elegant choice of basis functions; for instance, to move from the the PSF domain to OTF domain, one needs to change the basis functions

only. In other words, this process does not demand any extra calculation and the coefficients that have been calculated for PSF domain, can be used directly for other domains too.

# Chapter 4

# Depth of Field

We discuss the limit of depth of field extension for an imaging system using aspheric surfaces. In particular we consider a general imaging system in the sense that it has arbitrary pupil-function phase and present the trade-off between the depth of field of the system and the spectral signal-to-noise-ratio (SNR) over an extended depth of field. In doing so we use the relation between the conservation of ambiguity and modulation transfer function (MTF) on one hand and the relation between the spectral SNR and MTF on the other hand. Using this, we rigorously derive the expression for the tightest upper bound for the minimum spectral SNR, i.e. the limit of spectral SNR improvement. We also draw the relation between our result and the conservation of brightness theorem and establish that our result is the spectral version of the brightness conservation theorem.

## 4.1 Introduction

A common problem encountered in the design of imaging systems consists of finding the right balance between the light gathering ability and the depth of field (DOF) of the system. A common metric for the DOF is given by the Rayleigh criteria, defined as the range of deviation of the distance from the pupil plane to the object plane that gives rise to a quarter wave of defocus [35]. However, in many imaging systems it is more useful to define the DOF in terms of the spectral SNR defined as the ratio between the power spectra of the relevant signal present in the image over the noise

[8]. This is the case, for example, in task-based imaging systems in which a minimum SNR level is required within a specified range of spatial frequencies for a minimally required level of system performance (See Chapter two). Having high SNR at the detector of an imaging system over a large range of depths of field has been the utmost goal in many imaging system designs [8, 9, 10].

Traditionally, most of the attention in the literature has been focused on increasing the depth of field for special problems of interest. This typically includes cases of successfully designed imaging systems that work in an extended depth of field. Usually in these systems SNR is shown to be within acceptable limits depending on the particular goal. There are however cases in which a subclass of design problems (for instance, cubic-phase pupil function) are studied analytically where the limits of extension of depth of field in terms of generic acceptable SNR is discussed more rigorously [9, 13, 14]. Nevertheless these results[9, 13] are all, to some extent, for special cases and so far the literature has not explored the limit of extension of depth of field in general. Thus, the main question remains unanswered: what is the limit to which one can extend the depth of field and still keep the SNR above a required acceptance limit?

We start with a quick review of the depth of field extension methods. Traditionally (as we have all experienced with our professional cameras) one can extend the depth of field by controlling the aperture stop size. Albeit very simple, this method has serious drawbacks such as significantly reducing the optical power available and the highest spatial frequency [15]. This limits the practical use of this method to very short ranges of depth of field [16]. In 1995 E. R. Dowski and W. T. Cathey introduced a new method for extending the depth of field called wavefront coding [17]. Wavefront coding combines aspheric optical elements and digital signal processing to extend the depth of field of imaging systems. In general, wavefront coding can be thought of as an example of pupil function engineering. Because elements used in wavefront coding are typically non-absorbing phase elements, they allow the exposure and illumination to be maintained while producing the depth of field of a slower system [18, 19, 20]. Although numerous important industrial problems are solved using pupil function engineering, there is no concrete statement about the extent pupil function engineering can improve SNR over the depth of field of interest.

In this Chapter we introduce a relation between the spectral SNR and the MTF by analyzing imaging systems. Then we establish a relation between the MTF and the ambiguity function. Using these two relations and the concept of conservation of the ambiguity function, we derive a limit on the amount of spectral SNR available in an imaging system. This amount, which is bounded by a finite value, can be distributed over desired ranges of defocus. We show how our results can be used to establish a limit on SNR improvement in extended depth of field imaging systems.

In the next Section we introduce the spectral SNR and we derive its relation with the MTF. In Section 4.3 we introduce the relation between the ambiguity function and the modulation transfer function. In Section 4.4 we present the main result of this Chapter: The spectral SNR conservation law. In Section 4.5 we discuss the direct applications of our results. In particular, we present an upper bound for the minimum spectral SNR, thus finding the limit on SNR improvement in imaging systems with extended depth of field. In this Section, we also draw the relation between our results and the brightness conservation theorem and interpret our result as the spectral version of this theorem.

## 4.2 The Spectral Signal-to-Noise Ratio

We are interested in studying the extension of depth of field by optimizing the pupil-function phase. For simplicity, we consider a one-dimensional imaging system with no absorption and arbitrary phase at the pupil plane as shown in Fig. 4-1. Let $\mathcal{P}(\hat{x};\mathbf{c}) = \text{rect}(\hat{x})\exp[\mathbf{i}k\,w(\hat{x};\mathbf{c})]$ be the pupil function where $\hat{x}$ is the normalized Cartesian coordinate system in the pupil plane, $\mathbf{c}$ is the arbitrary vector containing all the parameters that we tune to increase the depth of field (hence arbitrary phase), $w$ is the wavefront at the pupil plane and rect is as defined in Ref. [35]. We call $\mathbf{c}$ the design vector. Also let $\widehat{W}_{20} = [D^2/(8\lambda)](1/d_o + 1/d_i - 1/f)$ be the dimensionless defocus which is normalized using the illumination wavelength. Note that $d_i$, $d_o$ and $f$ represent the image-plane to the pupil-plane distance, the object-plane to the pupil-plane distance, and the focal length of the imaging system, respectively. Also, under incoherent illumination, let us call the optical power leaving the object plane $\mathscr{I}_{obj}$. In addition, we consider white read noise in our imaging system. This means noise

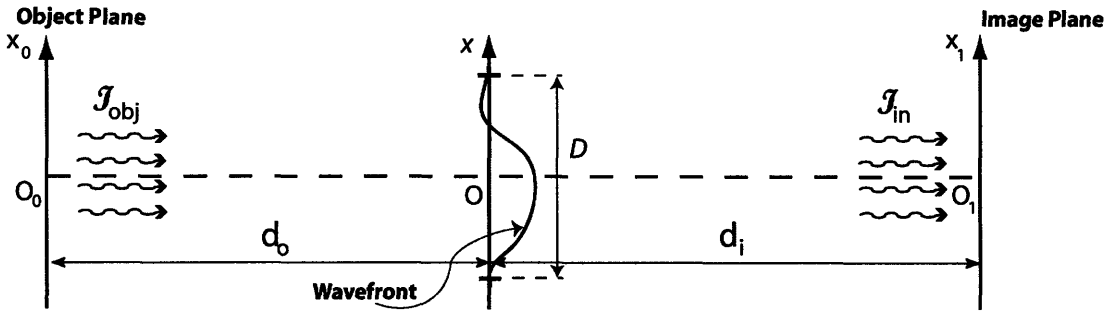



Figure 4-1: Schematic view of the imaging system under consideration. $O$ is the center of aperture, $O_0$ is the center of the object plane and $O_1$ is the center of the image plane. $\mathscr{I}_{obj}$ is the power leaving the object plane and $\mathscr{I}_{in}$ is the power arriving at the image plane. Finally, $D$ is the width of the aperture.

power will have no spectral information.

We define the spectral SNR, as the ratio of signal power spectrum to the noise power spectrum. This ratio has two parts: one that depends on the spatial frequency and another that does not. The latter one is called the carrier SNR. Here, we consider our predominant noise source to be read noise which is assumed to be spectrally white. Thus, the only spatial-frequency-dependent part of SNR comes from the signal power spectrum. Considering this definition, the spectral SNR for an arbitrary imaging system with an arbitrary object can be written as [8, 36]

$$SNR(u, \widehat{W}_{20}; \mathbf{c}) = \mathcal{SNR}_c \, \widehat{\mathscr{I}_{obj}}^{\,2}(u) \, MTF^2(u; \widehat{W}_{20}; \mathbf{c}), \tag{4.1}$$

where $\widehat{\mathscr{I}_{obj}}$ is the normalized power spectrum of the object. Note that both $\widehat{\mathscr{I}_{obj}}$ and MTF are normalized to one and are dimensionless. These two expressions govern the spectral dependence of SNR; here, $u$ represents the normalized spatial frequency. Comparison of the above equation with Eq. (15) of Ref. [8] shows that a factor of $\widehat{\mathscr{I}_{obj}}(u)$ [or $\mathcal{R}(u)$ with corresponding reference's notation] is missing in Ref. [8]. Also note that in special case of task-based imaging systems, Eq. (4.1) has another normalized term ($|S(u)|^2$) to account for the spatial frequencies of interest to the task at hand. However, in general, this term could be assumed to be spectrally flat and thus neglected.

Note that $\mathcal{SNR}_c$ represents the mean SNR present over the fraction of interest of the im-

age, averaged over all spatial frequencies. As such, it is not a function of normalized spatial frequency. Also note that, in the case of imaging systems with arbitrary pupil-function phase, we have $\partial \mathcal{SNR}_c / \partial \mathbf{c} = \mathbf{0}$. This is because, change of phase of pupil function does not change the absorbance of the imaging system. Note that for finite objects $\mathcal{SNR}_c$ is independent of $d_o$ [8]. Also, without loss of generality we assume that all defocus in our imaging system is due to change in $d_o$; thus, we conclude that $\mathcal{SNR}_c$ is also independent of defocus ($\partial \mathcal{SNR}_c / \partial \widehat{W}_{20} = 0$). Further details of the structure of the carrier SNR, $\mathcal{SNR}_c$, is covered in Section 4.5.1.

To summarize, we conclude that the dimensionless $\mathcal{SNR}_c$ is independent of $u$, $\mathbf{c}$ and $\widehat{W}_{20}$, but the spectral SNR is a function of square of MTF and square of normalized power spectrum of the object. Note that defining the spectral SNR allows to design and optimize the SNR of imaging system more specifically and based on the range of frequencies of interest.

## 4.3 The Ambiguity Function and the Modulation Transfer Function

In this Section, we introduce the ambiguity function and some of the fundamental results that has been previously developed. Furthermore, we establish the relation between the ambiguity function and the MTF and thus motivate how the fundamental results regarding ambiguity function can be applied to imaging system design and analysis.

Given a signal, $f(\tau)$, the ambiguity function, $A_f(\xi, \zeta)$, is defined as[37]

$$A_f(\xi, \zeta) = \int_{-\infty}^{\infty} f(\tau + \xi/2) f^*(\tau - \xi/2) \exp(-2\pi \mathbf{i}\, \zeta\, \tau)\, \mathrm{d}\tau, \tag{4.2}$$

where $\mathbf{i} = \sqrt{-1}$. Now we relate MTF to $A_{\mathcal{F}(\mathcal{P})}$, where $\mathcal{F}$ represents the Fourier transform operator and $\mathcal{P}$, as defined earlier, is the pupil function. It has been shown that[38, 29, 39]

$$\mathcal{H}(u; \widehat{W}_{20}; \mathbf{c}) = \frac{1}{\mathcal{A}} A_{\mathcal{F}(\mathcal{P})}(4u\widehat{W}_{20}, 2u; \mathbf{c}), \tag{4.3}$$

where $\mathcal{H}$ is the optical transfer function (OTF) of the imaging system and $\mathcal{A}$ is the area of aperture stop, which in our one dimensional case would be the length of the aperture stop. Using the relation between MTF and OTF [31, 35], we have

$$MTF(u;\widehat{W}_{20};\mathbf{c}) = \frac{1}{\mathcal{A}}\left|A_{\mathcal{F}(\mathcal{P})}(4u\widehat{W}_{20}, 2u;\mathbf{c})\right|. \tag{4.4}$$

Note that both the ambiguity function and the MTF are two-dimensional functions of normalized spatial frequency and defocus. In what follows we illustrate how $MTF(u;\widehat{W}_{20})$ and $\frac{1}{\mathcal{A}}\left|A_{\mathcal{F}(\mathcal{P})}(\mathcal{X},\mathcal{Y})\right|$ or equivalently $\frac{1}{\mathcal{A}}\left|A_{\mathcal{F}(\mathcal{P})}(r\angle\phi)\right|$ are related, where

$$\begin{aligned} \mathcal{X} &\equiv 4u\widehat{W}_{20} \\ \mathcal{Y} &\equiv 2u, \end{aligned} \tag{4.5}$$

$$\begin{aligned} r &\equiv 2|u|\sqrt{1+\cot^2(\phi)} \\ \phi &\equiv \text{arccot2}\,(4u\widehat{W}_{20}, 2u), \end{aligned} \tag{4.6}$$

where arccot2 $(x, y)$ is the angle between the $x$-axis and the line that connects origin and $(x, y)$ in the $x-y$ plane. Now, to span values of MTF for a particular spatial frequency over all values of defocus one needs to move along the horizontal line that intersects the $\mathcal{Y}$-axis at $\mathcal{Y} = 2u$, where $u$ is the normalized spatial frequency of interest. By the same token, to span values of MTF for a particular defocus over all values of spatial frequency one needs to move along the line $\mathcal{Y} = \mathcal{X}/(2\widehat{W}_{20})$. This line passes through the origin and makes angle $\phi$ with the $\mathcal{X}$-axis. It is crucial to understand that there is a one-to-one relation between each pair of $(u;\widehat{W}_{20})$ and $(\mathcal{X},\mathcal{Y})$ [or $(r\angle\phi)$], where $u, \widehat{W}_{20}, \mathcal{X}, \mathcal{Y} \in \Re$, $r \in [0, \infty)$ and $\phi \in [0, 2\pi)$. Also it should be noted that the pair of $(u;\widehat{W}_{20})$ neither makes a

Cartesian nor polar basis for MTF considering the way MTF is shown by the ambiguity function. Rather $(u; \widehat{W}_{20})$ is related to each of these descriptions through Eqs. (4.5) and (4.6).

Now, considering the ambiguity function $A_f(\xi, \zeta) \equiv A_f(\rho \angle \theta)$ and using the concept of conservation of ambiguity [40], we have

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |A_f(\xi,\zeta)|^2 \, \mathrm{d}\xi \, \mathrm{d}\zeta = \int_0^{2\pi}\int_0^{\infty} |A_f(\rho\angle\theta)|^2 \, \rho \, \mathrm{d}\rho \, \mathrm{d}\theta = [\mathcal{E}(f)]^2, \tag{4.7}$$

where $\mathcal{E}(f)$ stands for the energy of signal $f$ and is defined as

$$\mathcal{E}(f) = \int_{-\infty}^{\infty} f(\tau) f^*(\tau) \, \mathrm{d}\tau. \tag{4.8}$$

Now using Parseval's theorem [41] we have

$$\mathcal{E}(f) = \int_{-\infty}^{\infty} \underset{\tau \to T}{\mathcal{F}} \{f(\tau)\} \left[ \underset{\tau \to T}{\mathcal{F}} \{f(\tau)\} \right]^* \mathrm{d}T. \tag{4.9}$$

## 4.4 Conservation of the Spectral Signal-to-Noise Ratio

In this Section we introduce the concept of conservation of the spectral SNR. In particular we find the sum of all received power in the image plane as a spectrally flat object moves along the depth of field. We show that, this sum is finite except at zero spatial frequency. We also prove that, this sum does not change as the phase of the pupil function changes. This is what we refer to as the conservation of the spectral SNR. Finally we derive a similar result for the spectral SNR of an arbitrary object (as opposed to a spectrally flat object).

As opposed to an arbitrary object, we define a spectrally flat object as follows.

**Definition 4.4.1.** *A spectrally flat object is defined as any object whose reflected optical power are independent of the normalized spatial frequency.* $(\partial \widehat{\mathscr{I}}_{obj-i}/\partial u = 0)$.

It immediately follows that $\widehat{\mathscr{I}}_{obj-i}(u) = 1$. The main result of this Chapter is presented in the

following Theorem.

**Theorem 4.4.1.** *Let $SNR_i = SNR_i(u, \widehat{W}_{20}; \mathbf{c})$ be the spectral signal-to-noise ratio of an arbitrary imaging system with a spectrally flat object. Also let,*

$$C(u, \mathbf{c}) = \int_{-\infty}^{\infty} SNR_i(u, \widehat{W}_{20}; \mathbf{c})\, \mathrm{d}\widehat{W}_{20}.$$

*Then we have*

$$\frac{\partial C(u, \mathbf{c})}{\partial \mathbf{c}} = 0,$$

*and*

$$C(u; \mathbf{c}) = \mathcal{SNR}_c\, \frac{1 - |u|}{8\,|u|}\, \mathrm{rect}\left(\frac{u}{2}\right).$$

Theorem 4.4.1 states that the amount of spectral SNR for an arbitrary imaging system with a spectrally flat object over all values of defocus cannot be changed with the change of phase of pupil function. The following corollary immediately follows from Theorem 4.4.1.

**Corollary 4.4.1.** *Let $SNR = SNR(u, \widehat{W}_{20}; \mathbf{c})$ be the signal-to-noise ratio of an arbitrary imaging system with arbitrary object. Then we have*

$$\int_{-\infty}^{\infty} SNR(u, \widehat{W}_{20}; \mathbf{c})\, \mathrm{d}\widehat{W}_{20} \leq C(u, \mathbf{c}).$$

Before proving the above claims, we present the following Lemma which in itself reveals an interesting property of MTF of an arbitrary imaging system.

**Lemma 4.4.1.** *For the modulation transfer function of any arbitrary imaging system, we have*

$$\int_{-\infty}^{\infty} MTF^2(u; \widehat{W}_{20}; \mathbf{c})\, \mathrm{d}\widehat{W}_{20} = \frac{1 - |u|}{8\,|u|}\, \mathrm{rect}(\frac{u}{2}).$$

**Proof:**

We start with the left-hand side of the Lemma 4.4.1. Let us call the left-hand side $\Phi(u; \mathbf{c})$, then

for $u > 0$ we have

$$\begin{aligned}
\Phi(u;\mathbf{c}) &= \int_{-\infty}^{\infty} MTF^2(u;\widehat{W}_{20};\mathbf{c})\,\mathrm{d}\widehat{W}_{20} \qquad (4.10)\\
&= \frac{1}{\mathcal{A}^2}\int_{-\infty}^{\infty}\left|A_{\mathcal{F}(\mathcal{P})}(4u\widehat{W}_{20},2u;\mathbf{c})\right|^2\,\mathrm{d}\widehat{W}_{20}\\
&= \frac{1}{4u\mathcal{A}^2}\int_{-\infty}^{\infty}\left|A_{\mathcal{F}(\mathcal{P})}(\xi,\zeta;\mathbf{c})\right|^2\,\mathrm{d}\xi\\
&= \frac{1}{4u\mathcal{A}^2}\int_{-\infty}^{\infty}\left|\mathcal{F}_{\xi\to a}\left\{A_{\mathcal{F}(\mathcal{P})}(\xi,\zeta;\mathbf{c})\right\}\right|^2\,\mathrm{d}a\\
&= \frac{1}{4u\mathcal{A}^2}\int_{-\infty}^{\infty}\left|\mathcal{P}\left(a+\frac{\zeta}{2};\mathbf{c}\right)\mathcal{P}^*\left(a-\frac{\zeta}{2};\mathbf{c}\right)\right|^2\,\mathrm{d}a\\
&= \frac{1}{4u\mathcal{A}^2}\int_{-\infty}^{\infty}\mathrm{rect}\left(\frac{a}{2}+\frac{\zeta}{4}\right)\mathrm{rect}\left(\frac{a}{2}-\frac{\zeta}{4}\right)\,\mathrm{d}a\\
&= \frac{1}{2u\mathcal{A}^2}\left(1-|u|\right)\mathrm{rect}\left(\frac{u}{2}\right)\\
&= \frac{1-|u|}{8u}\,\mathrm{rect}\left(\frac{u}{2}\right).
\end{aligned}$$

The second equality follows from Eq. (4.4). The third equality follows from change of variable: $(\xi,\zeta) \equiv (4u\widehat{W}_{20}, 2u)$. The fourth equality follows from Parseval's Theorem. The fifth equality follows from the definition of the ambiguity function in Eq. (4.2). The sixth equality follows from the expression for the pupil function in Section 4.2. The seventh equality follows from performing the integration over $a$ and using $\zeta = 2u$. The last equality follows from the fact that we are working with one-dimensional and normalized Cartesian coordinate system and thus $\mathcal{A} = 2$. Using the same method it can be shown that for $u < 0$ we have

$$\Phi(u;\mathbf{c}) = \frac{1-|u|}{-8u}\,\mathrm{rect}\left(\frac{u}{2}\right). \qquad (4.11)$$

Considering the fact that $MTF(u=0;\widehat{W}_{20};\mathbf{c}) = 1$, it is clear that $\Phi(u=0;\mathbf{c}) = \infty$. Thus we have

$$\Phi(u;\mathbf{c}) = \frac{1-|u|}{8\,|u|}\,\mathrm{rect}\left(\frac{u}{2}\right), \qquad (4.12)$$

for all $u$.

□

**Proof of Theorem 4.4.1**: We directly evaluate $C(u, \mathbf{c})$ and show that it is independent of $\mathbf{c}$. Substituting Eq. (4.1) in the equation of $C(u, \mathbf{c})$, we have

$$C(u, \mathbf{c}) = \int_{-\infty}^{\infty} \mathcal{SNR}_c \, \widehat{\mathscr{I}_{obj-i}}^2(u) \, MTF^2(u; \widehat{W}_{20}; \mathbf{c}) \, \mathrm{d}\widehat{W}_{20}. \tag{4.13}$$

Note that $\mathcal{SNR}_c$ is independent of defocus and spatial frequency for an ideal imaging system. Furthermore $\widehat{\mathscr{I}_{obj-i}} = 1$ by definition. Thus we have

$$C(u, \mathbf{c}) = \mathcal{SNR}_c \int_{-\infty}^{\infty} MTF^2(u; \widehat{W}_{20}; \mathbf{c}) \, \mathrm{d}\widehat{W}_{20}. \tag{4.14}$$

Now using Lemma 4.4.1 we have

$$C(u; \mathbf{c}) = \mathcal{SNR}_c \, \frac{1 - |u|}{8\,|u|} \, \mathrm{rect}\left(\frac{u}{2}\right). \tag{4.15}$$

Theorem 4.4.1 immediately follows by considering the fact that $C$ is not a function of $\mathbf{c}$.

□

**Proof of Corollary 4.4.1:** Considering the fact that the spectral SNR is a power spectrum ratio and therefore, always positive; and, in general, $0 \leq \widehat{\mathscr{I}_{obj}} \leq 1 = \widehat{\mathscr{I}_{obj-i}}$, Corollary 4.4.1 immediately follows.

## 4.5 Discussion

In this Section we review the implications of our results. In Section 4.5.1 we derive the limit of extension of the depth of field using pupil function engineering. In doing so, we also discuss how this limit may be evaluated in practice. In particular we discuss the calculation of $\mathcal{SNR}_c$ and its main

components. In Section 4.5.2 we discuss the relation between our result and the fundamental theorem of conservation of the brightness. In particular, we establish that our result can be thought of as the spectral version of the brightness conservation theorem. We discuss how our result regarding the conservation of the spectral SNR can clarify some of the ambiguity of the brightness conservation theorem.

### 4.5.1 Limit of Extension of Depth of Field

The main result of this Section is shown in Eq. (4.16). Here, $\overline{SNR}(u)$ is the the tightest upper bound for minimum spectral SNR, i.e. the limit of spectral SNR improvement. Given, a particular problem specification, designers can easily and promptly evaluate the limit of SNR improvement for a particular depth of field of interest. Alternatively, designers can calculate the maximum depth of field possible, given a required minimum spectral SNR.

$$\overline{SNR}(u) = \frac{{E_0}^2 T^2}{D^2 + 4{d_i}^2} \left(\frac{\eta}{h\nu}\right)^2 \frac{{e_t}^2 {f_f}^2 p^4}{{m_{RN}}^2} \frac{1-|u|}{|u|} \operatorname{rect}\left(\frac{u}{2}\right) \frac{\lambda}{\Delta\frac{1}{d_o}}. \tag{4.16}$$

Theorem 4.4.1 and Corollary 4.4.1 state that all pupil function engineering does is to redefine the distribution of spectral SNR among different defocus values. This is because in practice, the total amount of available spectral SNR is limited; i.e. $C(u;\mathbf{c})$ is finite. To see this, we rewrite Eq. (4.15)

$$C(u) = \mathcal{SNR}_c \frac{1-|u|}{8\,|u|} \operatorname{rect}\left(\frac{u}{2}\right). \tag{4.17}$$

Note that, taking into account Theorem 4.4.1, we have only shown the functionality of $C$ with respect to spatial frequency, i.e. $C(u)$. References [8, 42, 43] provides a comprehensive treatment of carrier SNR, $\mathcal{SNR}_c$. Carrier SNR depends on illumination, imaging system absorbance and detector noise amongst other things,

$$\mathcal{SNR}_c = {E_0}^2 T^2 \frac{D^2}{D^2 + 4{d_i}^2} \left(\frac{\eta}{h\nu}\right)^2 \frac{{e_t}^2 {f_f}^2 p^4}{{m_{RN}}^2}, \tag{4.18}$$

where $E_0$ is the illumination irradiance, $T$ is the mean power transmission through the imaging system accounting for optical power loss (usually assumed to be unity), $e_t$ is the exposure time, $f_f$ is the detector array fill factor, $p$ is the pixel pitch and $m_{RN}$ is the number of read noise electrons. Now, substituting the expression for $\mathcal{SNR}_c$ in Eq. (4.17), we have

$$C(u) = {E_0}^2 T^2 \frac{D^2}{D^2 + 4{d_i}^2} \left(\frac{\eta}{h\nu}\right)^2 \frac{{e_t}^2 {f_f}^2 p^4}{{m_{RN}}^2} \frac{1 - |u|}{8\,|u|} \operatorname{rect}\left(\frac{u}{2}\right). \tag{4.19}$$

Calculating the actual value of $C(u)$ using Eq. (4.19) is straight forward. Note that Eq. (4.19) consists of three parts. The first part, ${E_0}^2 T^2 D^2/(D^2 + 4{d_i}^2)$, represents light collecting capacity of the imaging system. This part shows the power per area that enters the imaging system. The second part, $[\eta/(h\nu)]^2 ({e_t}^2 {f_f}^2 p^4)/({m_{RN}}^2)$, represents the detecting capacity of the imaging system. This part shows the amount of carrier SNR delivered by the imaging system per unit of power per area. The last part, $(1 - |u|)/(8\,|u|)$ rect $(u/2)$, represents the normalized spatial frequency dependence of $C(u)$. In other words, it shows the spectral behavior of $C(u)$.

One way to think of $C(u)$ is to consider it as the area under the plot of spectral SNR when it is plotted as a function of defocus. This description of $C(u)$ forms the basis for evaluating the limit on SNR improvement in imaging systems with extended depth of field.

The tightest upper bound for minimum spectral SNR, i.e. the limit of spectral SNR improvement, $\overline{SNR}(u)$, can be obtained by considering a rectangle whose length, width and area are the range of defocuses of interest (which is directly related to the range of depths of field of interest), $\overline{SNR}(u)$, and $C(u)$, respectively. Then, $\overline{SNR}(u)$ is

$$\overline{SNR}(u) = \frac{C(u)}{\Delta\widehat{W}_{20}}. \tag{4.20}$$

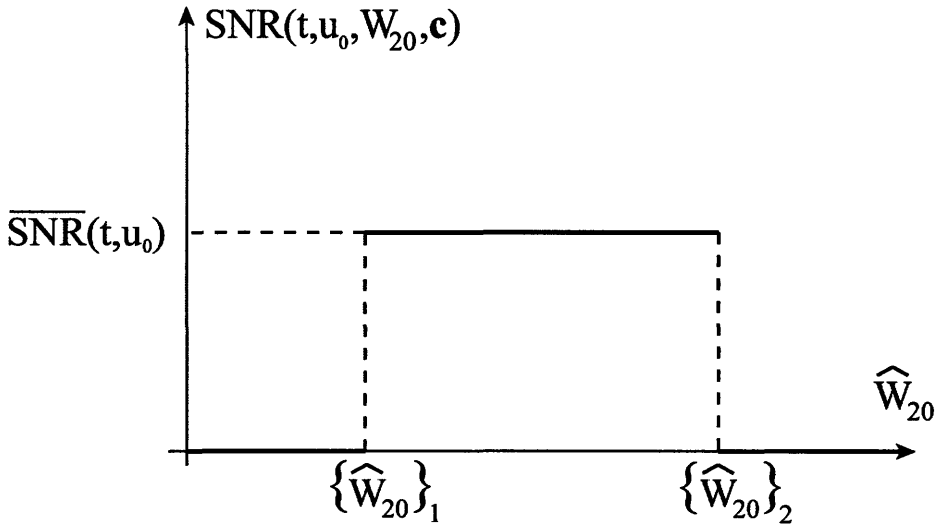



Figure 4-2: Plot of spectral SNR as a function of defocus.

See Figure 4-2 for details. Substituting the expressions for $\widehat{W}_{20}$ and $C(u)$ in Eq. (4.20), we have

$$\overline{SNR}(u) = \frac{{E_0}^2 T^2}{D^2 + 4{d_i}^2} \left(\frac{\eta}{h\nu}\right)^2 \frac{{e_t}^2 {f_f}^2 p^4}{{m_{RN}}^2} \frac{1-|u|}{|u|} \operatorname{rect}\left(\frac{u}{2}\right) \frac{\lambda}{\Delta\frac{1}{d_o}}, \tag{4.21}$$

which is the desired result.

### 4.5.2 The Brightness Conservation Theorem

Consider the imaging system shown in Fig. 4-1 and a spectrally flat object. Also assume this system only has white Gaussian noise. We can find the total transferable SNR from the object volume to the image plane by

$$\begin{aligned} T_{\mathtt{SNR}} &= \int \mathtt{SNR}(\widehat{W}_{20}) \mathrm{d}\widehat{W}_{20} \\ &= c_1 \int P_i(\widehat{W}_{20})\, A_i(\widehat{W}_{20})\, \Omega_i(\widehat{W}_{20})\, \mathrm{d}\widehat{W}_{20}, \end{aligned} \tag{4.22}$$

where SNR is the conventional signal-to-noise ratio (note the difference between SNR and $SNR(u)$, the spectral signal-to-noise ratio), $c_1$ is the proportionality constant, $P_i$ is the image brightness, $A_i$ is

the image area and $\Omega_i$ represents the solid angle. Index $i$ refers to the image when the object is at a location for which defocus is $\widehat{W}_{20}$. If we assume the object volume is produced by a one-dimensional object moving in the whole depth of field then the brightness of object remains constant. Using the conservation of the brightness theorem we conclude that image brightness in the above integral is constant and can be taken out of the integration (conservation of the brightness theorem tells us that the power per area per solid angle, i.e. brightness of object is equal in the object plane to the brightness of image in the image plane) [31]. Thus, we have

$$\begin{aligned} T_{\text{SNR}} &= c_1 P_i \int A_i(\widehat{W}_{20})\, \Omega_i(\widehat{W}_{20})\, \mathrm{d}\widehat{W}_{20} \\ &= \infty. \end{aligned} \tag{4.23}$$

where the second equality follows from simple geometric arguments. Equation 4.23 states that if a one-dimensional object moves from $d_o = 0$ to $d_o = \infty$, the summation of all measured power in the image plane is infinity. Also, clearly one can see that change of the phase of the pupil function neither changes this fact nor the amount of this summation.

Recalling the definition of $C(u)$ in Theorem 4.4.1, one notices that $C(u)$ looks at the transferable amount of SNR at a *particular spatial frequency* over the whole object volume. This clearly explains the physical meaning of $C(u)$. In particular, we are looking at the spectral behavior of $T_{\text{SNR}}$ which is the integral of brightness. As opposed to $T_{\text{SNR}}$ which is infinite, we have shown that $C(u)$ is finite except at $u = 0$ (See Theorem 4.4.1). In particular, we have shown

$$C(u) = \mathcal{SNR}_c \frac{1 - |u|}{8|u|} \text{rect}(u/2). \tag{4.24}$$

This means that even if the object moves from $d_o = 0$ to $d_o = \infty$ we cannot get infinite SNR at the image plane for a spatial frequency $u \neq 0$. Note that based on our discussion it is expected

$\int C(u)\mathrm{d}u = T_{\text{SNR}} = \infty$. This is explicitly shown below

$$\begin{aligned}
\int C(u)\mathrm{d}u &= \int \mathcal{SNR}_c \frac{1-|u|}{8|u|}\mathrm{rect}(u/2)\mathrm{d}u \qquad (4.25)\\
&= \mathcal{SNR}_c \int_{-1}^{1} \frac{1-|u|}{8|u|}\mathrm{d}u \\
&= \mathcal{SNR}_c \int_{0}^{1} \frac{1-u}{4u}\mathrm{d}u \\
&= \frac{\mathcal{SNR}_c}{4} \int_{0}^{1} \left(\frac{1}{u}-1\right)\mathrm{d}u \\
&= \frac{\mathcal{SNR}_c}{4} \left[\log(u)-u\right]_0^1 \\
&= \frac{\mathcal{SNR}_c}{4} \left\{[\log(1)-1] - \left[\lim_{u\to 0}\log(u) - 0\right]\right\} \\
&= \infty.
\end{aligned}$$

In other words it is interesting to note that by having an object go through the whole depth of field we can get infinite SNR at our image plane but the most *spectral* SNR we can get at the image plane under the same condition is limited by our result in Eq. 4.24. The former is a statement of the conservation of the brightness theorem and the latter, our result, can clearly be thought of as the spectral version of this theorem.

The other interesting implication is that the structure of $C(u)$ is independent of the pupil-function phase. This means we cannot design an imaging system (by changing the phase of the pupil function) which changes the fundamental spectral distribution of SNR of the object volume received at the image plane. This fact has been rigorously proved and can be thought of as the conservation of the spectral SNR.

# Chapter 5

# Conclusions

In this thesis, we studied pupil function engineering. In particular we developed new tools for this method; we solved some generic example problems and we answered a fundamental question in pupil function engineering.

In Chapter 2 we introduced a new analytic expression for the MTF. It was shown that this is an efficient expression that can approximate the MTF (assuming defocus and a cubic phase element at the pupil) with an analytic form. Also this approximation is adaptive in the sense that it can represent the average, lower bound or upper bound of the exact MTF. These adjustments can be easily performed by modifying the approximation kernel as described in App. B. The analytic expression for MTF significantly reduces the calculation time of MTF.

Using this expression, the optimal value for the cubic phase pupil function of two common problems in imaging was found. These two problems are uniform quality imaging (e.g. photography), and task-based imaging (e.g. iris recognition). The design of the cubic-phase pupil function using the approximate analytic optimum expressions and the problem specifications are illustrated. The results for both cases were compared through sample problems and their differences and similarities were discussed. The approximate analytic optimal solutions presented facilitate the process of pupil function engineering.

It was shown that the uniform quality imaging and task-based imaging problems are fundamen-

tally different. In the uniform quality imaging the MTF is almost symmetric around the plane of best focus. As the object reaches either end of the depth of field of interest, the MTF reaches its equal minimum at either of these two ends. However in task-based imaging the MTF is neither symmetric around the best focus, nor do we have the highest MTF at the original best focus. In fact in this case refocusing has removed the symmetry so that we have equal MTF at both ends of the desired depth of field *at the maximum spatial frequency of interest.* This shows how the system in each case has been optimized to do the particular job of interest.

In Chapter 3 we introduced a new method for analyzing the diffraction integral and evaluating the PSF. The new method is based on the use of higher order Airy functions along with a novel use of Zernike and Taylor expansions. This method is applicable when we are considering several aberrations and large defocus simultaneously. We have shown rigorously and verified by numerical simulations that the complexity of our expansion is independent of defocus and that it is stable in all ranges of defocus. The efficiency of the method compared to traditional ones has also been investigated and it is shown that the method not only does extremely faster than its alternates but also requires computational time that is independent of defocus.

The use of higher order Airy functions plays a key role in capturing the effect of different values of defocus in a simple expression whose complexity is independent of defocus. It was also shown in Theorem 3.5.1 that any arbitrary accuracy in any arbitrary region of interest could be achieved by a finite number of terms in the approximate function (Eq. (3.34)). Possible future work in this direction would be expanding our method for a nonuniform pupil transmittance function.

The complexity of this expansion is also invariant to resolution. Specifically, the time required for evaluating the PSF does not increase as the desired resolution increases. This could be a potential solution to some of the current problems in biological microscopy [32] and lithography [33] where having a high resolution information of PSF is critical. By providing an analytical solution for the diffraction integral, this approach, among other things, may also facilitate the process of multi-domain optimization, where the optical system and post-processing system are optimized together to increase the performance and/or reduce the cost of imaging systems. Aberration retrieval using

our PSF representation can be another possible future application of our analytical expression for PSF. This analytical expression for PSF may also help developing analytic treatment of incoherent imaging systems.

In Chapter 4 we have found an upper bound for the average achievable SNR in an imaging system with an extended depth of field. We have established that pupil function engineering cannot change this limit. In fact pupil function engineering can take advantage of the available resources (i.e. $C(t,u)$, the area under spectral SNR plot) to extend the depth of field of imaging systems to the maximum extent possible.

We established the relation between the MTF and the ambiguity function and showed that the ambiguity function is neither a Cartesian nor a polar representation of the MTF. We also introduced the relation between the spectral SNR and the square of the MTF and the square of object power spectrum. We showed that the spectral SNR is proportional to the other two. We also introduced the novel laws of conservation of MTF and conservation of spectral SNR.

Using our result, given a particular problem specification, designers can easily and promptly evaluate the limit of SNR improvement for a particular depth of field of interest. Alternatively, designers can calculate the maximum depth of field possible, given a required minimum spectral SNR. Hence we provided an upper bound on how well an imaging system with pupil function engineering can perform.

Finally we go over some of the possible future directions in pupil function engineering:

- The MTF approximation skim presented in this thesis can be easily generalized to other classes of pupil functions. This can be done by modifying the core approximation in App. B. This appraoch can be readily used to calculated the MTF of more complicated pupil functions, but to derive the corresponding accuracy results is one of the open problems.

- The PSF approximation method that is presented in Chapter 3 has a complexity which is independent of defocus. The next step in this direction is to extend our PSF representation to have a complexity which is independent of third order aberrations.

- One of other possible future directions is to perform analytic Multi-Domain Optimization (MDO). MDO referes to the optimization of the imaging system and the post processing algorithm simultaneously. Due to the lack of analytic expression for the MTF and/or the PSF, MDO has been always done numerically. However, using our analytic expression for the MTF and the PSF, one can perform analytic MDO.

# Appendix A

# Derivation of Eq. (2.1)

In this Appendix we derive an expression for the MTF of an imaging system with a cubic phase element installed in its aperture. We assume an imaging system with circular aperture being illuminated with incoherent light. Figure 2-1 shows a schematic view of our optical system. It can be shown that MTF is the normalized autocorrelation of the pupil function [35]. Using this general equation for MTF we have

$$MTF^e(f_x, f_y) = \left| \frac{\iint_{-\infty}^{\infty} P(x + \frac{\lambda d_i f_x}{2}, y + \frac{\lambda d_i f_y}{2}) P^*(x - \frac{\lambda d_i f_x}{2}, y - \frac{\lambda d_i f_y}{2}) \, \mathrm{d}x \, \mathrm{d}y}{\iint_{-\infty}^{\infty} P(x, y) P^*(x, y) \, \mathrm{d}x \, \mathrm{d}y} \right|, \tag{A.1}$$

where $MTF^e$ is the exact value of MTF, $P$ is the pupil function, $x$ and $y$ are Cartesian coordinates of the pupil and $f_x$ and $f_y$ are spatial frequencies in $x$ and $y$ directions. Before going further with Eq. (A.1), we first normalize it with respect to spatial coordinates, by dividing the coordinates by $D/2$, the aperture radius. Thus we have

$$MTF^e(u, v) = \left| \frac{\iint_{-\infty}^{\infty} \mathcal{P}(\hat{x} + u, \hat{y} + v) \mathcal{P}^*(\hat{x} - u, \hat{y} - v) \, \mathrm{d}\hat{x} \, \mathrm{d}\hat{y}}{\iint_{-\infty}^{\infty} \mathcal{P}(\hat{x}, \hat{y}) \mathcal{P}^*(\hat{x}, \hat{y}) \, \mathrm{d}\hat{x} \, \mathrm{d}\hat{y}} \right|, \tag{A.2}$$

where $\mathcal{P}$ is the pupil function with normalized variables, $\hat{x} = 2x/D$ and $\hat{y} = 2y/D$ are normalized Cartesian coordinates of the pupil and $u = \frac{\lambda d_i f_x}{D}$ and $v = \frac{\lambda d_i f_y}{D}$ are normalized spatial frequencies in $x$ and $y$ directions respectively. Before further analyzing Eq. (A.2), we formally introduce the cubic phase element. The cubic phase element is defined as

$$\Phi(\hat{x}, \hat{y}) = \alpha \left[ (\hat{x} + \delta)^3 + (\hat{y} + \delta)^3 \right]. \tag{A.3}$$

Here $\alpha$ is the cubic phase coefficient and $\delta$ is the cubic phase shift. These two quantities are in fact the design parameters of the cubic phase element as will be shown in Sections 2.3 and 2.4. Note that $\alpha$ and $\Phi$ in Eq. (A.3) have the dimension of length; however, $\hat{x}$, $\hat{y}$ and $\delta$ are dimensionless. Considering this definition, our pupil function, $\mathcal{P}$, would be

$$\mathcal{P}(\hat{x}, \hat{y}) = \exp\left\{ ki \left[ \Phi(\hat{x}, \hat{y}) + W_{20} \left( \hat{x}^2 + \hat{y}^2 \right) \right] \right\} \mathrm{circ}(\hat{x}, \hat{y}), \tag{A.4}$$

where circ function is defined as

$$\mathrm{circ}(\hat{x}, \hat{y}) = \begin{cases} 1 & \text{if } \hat{x}^2 + \hat{y}^2 \leq 1, \\ 0 & \text{otherwise,} \end{cases} \tag{A.5}$$

and $W_{20}$ is the defocus coefficient, defined as [31]

$$W_{20} = \frac{D^2}{8} \left( \frac{1}{d_i} + \frac{1}{d_o} - \frac{1}{f} \right), \tag{A.6}$$

where $k = 2\pi n/\lambda$, $f$, $d_i$, $d_o$ and $D$ are the wave number, imaging system focal length, distance from the image plane to the exit pupil, distance from the object plane to the entrance pupil and aperture

diameter respectively. The last three parameter definitions are illustrated in Fig. 2-1. Note that $W_{20}$ has the dimension of length.

Now, we focus on the performance of the MTF on two orthogonal axes. Note that the ultimate goal is to maximize MTF. However there is a fundamental limit to that due to conservation of ambiguity function [27, 28, 29]. A simple way of looking at this limit, is to consider the area under $(MTF^e)^2$ surface for all ranges of depth of field. The conservation of ambiguity tells us that this area is constant. In other words as we maximize MTF over some ranges of depth of field, we will reduce MTF elsewhere. This concept is covered in more details in Chapter 4. It has been observed that generally the most efficient way of managing this limit is to maximize MTF only on axis, thus keeping the used portion of this fixed area as small as possible. This way we can maximize MTF over significantly larger depths of field [27, 21, 18, 20, 24]. Thus, one needs to focus on maximizing the value of MTF along two orthogonal axes. Due to the symmetry of the problem, it suffices to analyze MTF in any of these two orthogonal directions. Thus by substituting $\mathcal{P}$ from Eq. (A.4) into Eq. (A.2) and setting $v$ equal to zero (i.e. looking at $u$-axis) we have

$$
\begin{aligned}
MTF^e(u,0) &= \left| \frac{\iint_{-\infty}^{\infty} \mathcal{P}(\hat{x}+u,\hat{y})\mathcal{P}^*(\hat{x}-u,\hat{y})\,\mathrm{d}\hat{x}\,\mathrm{d}\hat{y}}{\iint_{-\infty}^{\infty} \mathcal{P}(\hat{x},\hat{y})\mathcal{P}^*(\hat{x},\hat{y})\,\mathrm{d}\hat{x}\,\mathrm{d}\hat{y}} \right|, \qquad \text{(A.7)} \\
&= \left| \frac{1}{\iint_{-\infty}^{\infty} \mathrm{circ}(\hat{x},\hat{y})\mathrm{circ}^*(\hat{x},\hat{y})\,\mathrm{d}\hat{x}\,\mathrm{d}\hat{y}} \iint_{-\infty}^{\infty} \mathrm{e}^{ki\left\{\Phi(\hat{x}+u,\hat{y})+W_{20}\left[(\hat{x}+u)^2+\hat{y}^2\right]\right\}} \times \right. \\
&\quad \left. \mathrm{circ}(\hat{x}+u,\hat{y})\mathrm{e}^{-ki\left\{\Phi(\hat{x}-u,\hat{y})+W_{20}\left[(\hat{x}-u)^2+\hat{y}^2\right]\right\}} \mathrm{circ}^*(\hat{x}-u,\hat{y})\,\mathrm{d}\hat{x}\,\mathrm{d}\hat{y} \right|, \\
&= \frac{1}{\pi} \left| \mathrm{e}^{ki\left(6\alpha\delta^2 u+2\alpha u^3\right)} \times \right. \\
&\quad \left. \iint_{-\infty}^{\infty} \mathrm{e}^{ki\left[(12\alpha\delta u+4uW_{20})\,\hat{x}+(6\alpha u)\,\hat{x}^2\right]} \mathrm{circ}(\hat{x}+u,\hat{y})\,\mathrm{circ}(\hat{x}-u,\hat{y})\,\mathrm{d}\hat{x}\,\mathrm{d}\hat{y} \right|, \\
&= \frac{1}{\pi} \left| \iint_{-\infty}^{\infty} \mathrm{e}^{ki\left[(12\alpha\delta u+4uW_{20})\,\hat{x}+(6\alpha u)\,\hat{x}^2\right]} \mathrm{circ}(\hat{x}+u,\hat{y})\,\mathrm{circ}(\hat{x}-u,\hat{y})\,\mathrm{d}\hat{x}\,\mathrm{d}\hat{y} \right|.
\end{aligned}
$$

Now, using the definition of circ function in Eq. (A.5), we can rewrite the expression for the on-axis MTF as

$$MTF^e(u,0) = \frac{1}{\pi}\left|\int_{-y_m}^{y_m}\int_{-x_m}^{x_m} \mathrm{e}^{ki\left[(12\alpha\delta u + 4uW_{20})\,\hat{x} + (6\alpha u)\,\hat{x}^2\right]}\,\mathrm{d}\hat{x}\,\mathrm{d}\hat{y}\right|, \tag{A.8}$$

where $x_m$ and $y_m$ are defined as

$$\begin{aligned} x_m &= \sqrt{1-\hat{y}^2} - u, \\ y_m &= \sqrt{1-u^2}. \end{aligned} \tag{A.9}$$

Note that in Eq. (A.8), $12\alpha\delta$ and $4W_{20}$ have the same effect. Thus, without loss of generality we can assume $\delta$, the cubic phase shift, is equal to zero. This is because $W_{20}$ could be arbitrarily set (using either of $d_i$ or $f$) to make up for $\delta$. Another way of seeing this is to expand Eq. (A.3)

$$\Phi(\hat{x},\hat{y}) = \alpha\left[(\hat{x}^3+\hat{y}^3) + 3\delta(\hat{x}^2+\hat{y}^2) + 3\delta^2(\hat{x}+\hat{y}) + 6\delta^3\right]. \tag{A.10}$$

The forth term in Eq. (A.10) is constant phase shift and the third term is the tilt; both of which can be ignored. The second term is quadratic in pupil coordinates; hence has the same form as defocus and it can be ignored by refocusing; i.e. resetting $W_{20}$ by changing either $d_i$ or $f$. Thus, without loss of generality $\delta$ can be assumed to be zero in Eq. A.10. Hence we reach the desired result:

$$MTF^e(u,0) = \frac{1}{\pi}\left|\int_{-y_m}^{y_m}\int_{-x_m}^{x_m} \mathrm{e}^{ki\left[(4W_{20}u)\,\hat{x} + (6\alpha u)\,\hat{x}^2\right]}\,\mathrm{d}\hat{x}\,\mathrm{d}\hat{y}\right|. \tag{A.11}$$

# Appendix B

# The $MTF^{a1}$ Approximation

In this Appendix we derive the approximation $MTF^{a1}$ and present some results regarding its accuracy. In particular, we prove the first two parts of Theorem 2.2.1. To derive the approximation $MTF^{a1}$ we start with Eq. 2.1. Now, one can transform the argument of the exponential function in Eq. (2.1) to a complete square. Note that since we are interested in the absolute value, adding and subtracting constant values to or from the phase will not change the result. Doing that along with a change of variable as shown in Eq. (B.2) leads to the following expression

$$\begin{aligned} MTF^{e}(u,0) &= \frac{1}{\pi}\left|\int_{-y_m}^{y_m}\int_{-x_m}^{x_m}\exp(\mathrm{i}\,X^2)\,\mathrm{d}\hat{x}\,\mathrm{d}\hat{y}\right|, \\ &= \frac{1}{\pi\sqrt{6ku\alpha}}\left|\int_{-y_m}^{y_m}\int_{X_1}^{X_2}\exp(\mathrm{i}\,X^2)\,\mathrm{d}X\,\mathrm{d}\hat{y}\right|, \\ &= \frac{1}{\pi\sqrt{6ku\alpha}}\left|\int_{-y_m}^{y_m}\left[\int_{0}^{X_2}\exp\left(\mathrm{i}\,X^2\right)\mathrm{d}X-\int_{0}^{X_1}\exp\left(\mathrm{i}\,X^2\right)\mathrm{d}X\right]\,\mathrm{d}\hat{y}\right|, \end{aligned} \tag{B.1}$$

where

$$
\begin{aligned}
X &\equiv \sqrt{6ku\alpha}\,\hat{x} + \sqrt{\frac{2ku}{3\alpha}}W_{20}, \\
X_1 &\equiv -\sqrt{6ku\alpha}\,x_m + \sqrt{\frac{2ku}{3\alpha}}W_{20}, \\
X_2 &\equiv \sqrt{6ku\alpha}\,x_m + \sqrt{\frac{2ku}{3\alpha}}W_{20}.
\end{aligned}
\tag{B.2}
$$

At this point we have two integrals which can not be analytically evaluated in a closed form. However, we can use the definition of the following special function, the error function, to simplify one of the integrals.

$$
\int_0^X \exp(\mathbf{i}t^2)\mathrm{d}t = \sqrt{\frac{\pi}{2}}\,\frac{\mathbf{i}+1}{2}\,\mathrm{Erf}\left[\frac{(1-\mathbf{i})X}{\sqrt{2}}\right], \tag{B.3}
$$

To proceed further, we now introduce our novel approximation:

$$
\mathrm{Erf}\left[\frac{(1-\mathbf{i})X}{\sqrt{2}}\right] \approx \frac{2}{\pi}\,\mathrm{Arctan}\left(\sqrt{\pi}\,X\right), \tag{B.4}
$$

or

$$
\int_0^X \exp(\mathbf{i}t^2)\mathrm{d}t \approx \frac{\mathbf{i}+1}{\sqrt{2\pi}}\,\mathrm{Arctan}\left(\sqrt{\pi}\,X\right), \tag{B.5}
$$

Substituting this in Eq. (B.1), we get

$$MTF^{a1}(u,0) = \frac{1}{\pi\sqrt{6ku\alpha}}\left|\int_{-y_m}^{y_m}\frac{\mathbf{i}+1}{\sqrt{2\pi}}\left[\text{Arctan}\left(\sqrt{\pi}X_2\right)-\text{Arctan}\left(\sqrt{\pi}X_1\right)\right]\mathrm{d}\hat{y}\right| \quad \text{(B.6)}$$
$$= \frac{1}{\pi\sqrt{6ku\alpha\pi}}\left|\int_{-y_m}^{y_m}\left[\text{Arctan}\left(\sqrt{\pi}X_2\right)-\text{Arctan}\left(\sqrt{\pi}X_1\right)\right]\mathrm{d}\hat{y}\right|,$$

where $MTF^{a1}$ is the first approximation (a1) of the exact MTF ($MTF^{e}$). Substituting the values of $X_2$ and $X_1$ from Eqs. (B.2) and (2.2) to Eq. (B.6), we have

$$MTF^{a1}(u,0) = \frac{1}{\pi\sqrt{6\pi ku\alpha}}\int_{-y_m}^{y_m}\{ \quad \text{(B.7)}$$
$$\text{Arctan}\left[\sqrt{\frac{2\pi ku}{3\alpha}}(W_{20}-3u\alpha+3\alpha\sqrt{1-\hat{y}^2})\right]$$
$$-\ \text{Arctan}\left[\sqrt{\frac{2\pi ku}{3\alpha}}(W_{20}+3u\alpha-3\alpha\sqrt{1-\hat{y}^2})\right]\}\mathrm{d}\hat{y}.$$

Thus we need to solve the fundamental integral shown in Eq. (B.8).

$$I=\int \text{Arctan}\left(a+b\sqrt{1-\hat{y}^2}\right)\mathrm{d}\hat{y}, \quad \text{(B.8)}$$

where $a$ and $b$ are dummy constants. To solve the integral in Eq. (B.8) is an straightforward exercise in calculus and we present the result without going through the details. We have

$$I = \hat{y}\text{Arctan}\left(a+b\sqrt{1-\hat{y}^2}\right)-\frac{1}{b}\text{Arcsin}\left(\hat{y}\right)+\frac{\mathbf{i}}{2b}\left\{\sqrt{(a+\mathbf{i})^2-b^2}\times \quad \text{(B.9)}\right.$$
$$\left[\text{Arctan}\left(\frac{b\hat{y}}{\sqrt{(a+\mathbf{i})^2-b^2}}\right)-\text{Arctan}\left(\frac{(a+\mathbf{i})\hat{y}}{\sqrt{((a+\mathbf{i})^2-b^2)(1-\hat{y}^2)}}\right)\right]$$
$$-\sqrt{(a-\mathbf{i})^2-b^2}\times$$
$$\left.\left[\text{Arctan}\left(\frac{b\hat{y}}{\sqrt{(a-\mathbf{i})^2-b^2}}\right)-\text{Arctan}\left(\frac{(a-\mathbf{i})\hat{y}}{\sqrt{((a-\mathbf{i})^2-b^2)(1-\hat{y}^2)}}\right)\right]\right\}.$$

Equation (B.9) along with Eqs. (B.7) and (2.2) can be used to get the general solution for the

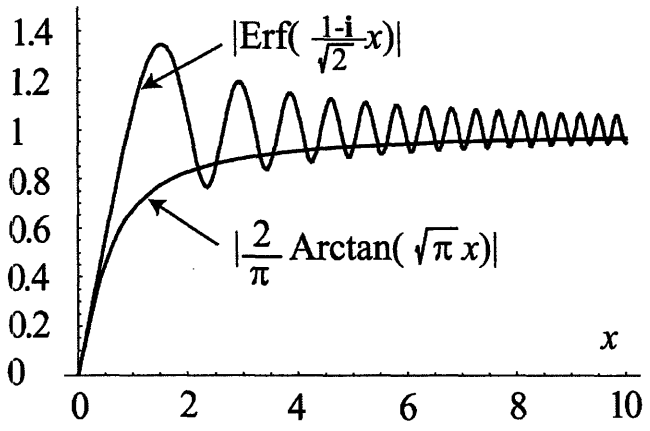



Figure B-1: Comparison of the imaginary error function and its approximation.

$MTF^{a1}$. In fact one can get the expression for the general $MTF^{a1}$ by simply replacing $a$ and $b$ by their corresponding values and applying the integral limits. Note that although the imaginary number ($\mathbf{i}$) exists in the expression of the equation, the outcome of $I$ is always a real number. This is because $[\mathbf{i}\,(x - x^*)] \in \Re$ where $x$ is arbitrary and $x^*$ is the complex conjugate of $x$.

To get some intuition regarding the approximation in Eqs. (B.4) and (B.5) one can see Fig. B-1. This figure shows the plot of Erf function. As it can be seen from this figure, Erf function has a highly oscillatory behavior as its argument gets larger than unity. Since in the end we are interested in the value of Eq. (B.1) which is the integral of the Erf function, we can expect the oscillation to have little effect in the final result, provided the range of the integration is large enough. This motivates us to use an approximation of the Erf function which possibly goes through the center of the oscillation of the Erf function. In particular one can see

$$\begin{aligned} \lim_{X\to\infty} \mathrm{Erf}\left[\frac{(1-\mathbf{i})X}{\sqrt{2}}\right] &= \lim_{X\to\infty} \frac{2}{\pi}\,\mathrm{Arctan}\left(\sqrt{\pi}\,X\right) = 1, \\ \left|\frac{\mathrm{d}}{\mathrm{d}X}\left\{\mathrm{Erf}\left[\frac{(1-\mathbf{i})X}{\sqrt{2}}\right]\right\}\right|_{X=0} &= \left|\frac{\mathrm{d}}{\mathrm{d}X}\left[\frac{2}{\pi}\,\mathrm{Arctan}\left(\sqrt{\pi}\,X\right)\right]\right|_{X=0} = \frac{2}{\sqrt{\pi}}. \end{aligned} \tag{B.10}$$

The other important factor in choosing an approximation is the feasibility of deriving a closed form expression for the second integral over $\hat{y}$. Note that by using this approximation [Eq. (B.4)],

we can easily proceed and evaluate the second integral over $\hat{y}$. Observe that, this was not possible without this approximation. Also the accuracy of this approximation is promising in the ranges of interest in imaging system design. We now discuss the latter issue. We remind the reader that $MTF^e$ is the exact value of the modulation transfer function (MTF) and $MTF^{a1}$ is the approximate version of that. We begin with the precise definition of MTF. In our definition we consider three main factors: the defocus, the cubic phase element and the aperture. In case of a circular aperture (with normalized radius), the expression of MTF along the u-axis is given by:

$$MTF^e(u,0) = \frac{1}{\pi}\left|\int_{-y_m}^{y_m}\int_{-x_m}^{x_m} e^{ki\left[(4W_{20}u)\,\hat{x}+(6\alpha u)\,\hat{x}^2\right]}\,d\hat{x}\,d\hat{y}\right|, \tag{B.11}$$

where

$$\begin{aligned} x_m &= \sqrt{1-\hat{y}^2}-u, \\ y_m &= \sqrt{1-u^2}. \end{aligned} \tag{B.12}$$

In case of a square aperture (which circumscribes the above normalized radius circle), the the expression of MTF along the u-axis is given by:

$$MTF^e(u,0) = \frac{1}{4}\left|\int_{-y_m}^{y_m}\int_{-x_m}^{x_m} e^{ki\left[(4W_{20}u)\,\hat{x}+(6\alpha u)\,\hat{x}^2\right]}\,d\hat{x}\,d\hat{y}\right|, \tag{B.13}$$

where

$$\begin{aligned} x_m &= 1-u, \\ y_m &= 1. \end{aligned} \tag{B.14}$$

We derive the accuracy expression for our approximation in case of square aperture. This choice

is solely due to simplicity of the square aperture case. Equation (B.13) can be simplified to

$$\begin{aligned} MTF^e(u,0) &= \frac{1}{4\sqrt{6ku\alpha}}\left|\int_{-y_m}^{y_m}\int_{X_1}^{X_2}\exp(\mathrm{i}\,X^2)\,\mathrm{d}X\,\mathrm{d}\hat{y}\right|, \\ &= \frac{1}{4\sqrt{6ku\alpha}}\left|\int_{-y_m}^{y_m}\left\{\int_0^{X_2}\exp(\mathrm{i}\,X^2)\,\mathrm{d}X-\int_0^{X_1}\exp(\mathrm{i}\,X^2)\,\mathrm{d}X\right\}\mathrm{d}\hat{y}\right|, \end{aligned} \tag{B.15}$$

where

$$\begin{aligned} X &\equiv \sqrt{6ku\alpha}\,\hat{x}+\sqrt{\frac{2ku}{3\alpha}}W_{20}, \\ X_1 &\equiv -\sqrt{6ku\alpha}\,x_m+\sqrt{\frac{2ku}{3\alpha}}W_{20}, \\ X_2 &\equiv \sqrt{6ku\alpha}\,x_m+\sqrt{\frac{2ku}{3\alpha}}W_{20}. \end{aligned} \tag{B.16}$$

Using Eqs. (B.14) and (B.16) we can simplify Eq. (B.15) to get

$$\begin{aligned} MTF^e(u,0) &= \frac{1}{2\sqrt{6ku\alpha}}\left|\int_0^{X_2}\exp(\mathrm{i}\,X^2)\,\mathrm{d}X-\int_0^{X_1}\exp(\mathrm{i}\,X^2)\,\mathrm{d}X\right| \\ &= \frac{1}{2\sqrt{6ku\alpha}}\left|\int_{X_1}^{X_2}\sum_{n=0}^{\infty}\frac{(\mathrm{i}\,X^2)^n}{n!}\,\mathrm{d}X\right| \\ &= \frac{1}{2\sqrt{6ku\alpha}}\left|\sum_{n=0}^{\infty}\int_{X_1}^{X_2}\frac{(\mathrm{i}\,X^2)^n}{n!}\,\mathrm{d}X\right| \\ &= \frac{1}{2\sqrt{6ku\alpha}}\left|\sum_{n=0}^{\infty}\left[\frac{\mathrm{i}^n\,{X_2}^{2n+1}}{(2n+1)n!}-\frac{\mathrm{i}^n\,{X_1}^{2n+1}}{(2n+1)n!}\right]\right|. \end{aligned} \tag{B.17}$$

Using the proposed approximation in Eq. (2.3) to substitute the integral over $X$ in Eq. (B.15) with its Arctan approximation, one can show that the approximate version of MTF is given by

$$
\begin{aligned}
MTF^{a1}(u,0) &= \frac{1}{4\sqrt{6ku\alpha}}\left|\int_{-y_m}^{y_m}\frac{\mathbf{i}+1}{\sqrt{2\pi}}\left\{\text{Arctan}\left(\sqrt{\pi}X_2\right)-\text{Arctan}\left(\sqrt{\pi}X_1\right)\right\}\mathrm{d}\hat{y}\right| \qquad \text{(B.18)}\\
&= \frac{1}{2\sqrt{6ku\alpha}}\left|\frac{\mathbf{i}+1}{\sqrt{2\pi}}\left\{\text{Arctan}\left(\sqrt{\pi}X_2\right)-\text{Arctan}\left(\sqrt{\pi}X_1\right)\right\}\right|,
\end{aligned}
$$

where the second equality follows from Eqs. (B.14) and (B.16).

In general as it was stated above we are interested to find the smallest bound for $\Delta$ given by

$$
\Delta \equiv \left|MTF^{e}(u,0)-MTF^{a1}(u,0)\right|. \qquad \text{(B.19)}
$$

To reach this goal, we start by using Eqs. (B.17) and (B.18) to get

$$
\begin{aligned}
\Delta &= \frac{1}{2\sqrt{6ku\alpha}}\left|\left|\sum_{n=0}^{\infty}\left[\frac{\mathbf{i}^n {X_2}^{2n+1}}{(2n+1)n!}-\frac{\mathbf{i}^n {X_1}^{2n+1}}{(2n+1)n!}\right]\right| \right. \qquad \text{(B.20)}\\
&\quad \left. -\left|\frac{\mathbf{i}+1}{\sqrt{2\pi}}\left\{\text{Arctan}\left(\sqrt{\pi}X_2\right)-\text{Arctan}\left(\sqrt{\pi}X_1\right)\right\}\right|\right|\\
&\leq \frac{1}{2\sqrt{6ku\alpha}}\left|\sum_{n=0}^{\infty}\left[\frac{\mathbf{i}^n {X_2}^{2n+1}}{(2n+1)n!}-\frac{\mathbf{i}^n {X_1}^{2n+1}}{(2n+1)n!}\right]\right.\\
&\quad \left. -\frac{\mathbf{i}+1}{\sqrt{2\pi}}\left\{\text{Arctan}\left(\sqrt{\pi}X_2\right)-\text{Arctan}\left(\sqrt{\pi}X_1\right)\right\}\right|\\
&\leq \frac{1}{2\sqrt{6ku\alpha}}\left\{\left|\sum_{n=0}^{\infty}\left[\frac{\mathbf{i}^n {X_2}^{2n+1}}{(2n+1)n!}\right]-\frac{\mathbf{i}+1}{\sqrt{2\pi}}\text{Arctan}\left(\sqrt{\pi}X_2\right)\right|\right.\\
&\quad \left. +\left|\sum_{n=0}^{\infty}\left[\frac{\mathbf{i}^n {X_1}^{2n+1}}{(2n+1)n!}\right]-\frac{\mathbf{i}+1}{\sqrt{2\pi}}\text{Arctan}\left(\sqrt{\pi}X_1\right)\right|\right\}\\
&\leq \frac{1}{2\sqrt{6ku\alpha}}\left[Er(X_2)+Er(X_1)\right],
\end{aligned}
$$

where the first equality follows from the definition. The first and second inequalities follow from the properties of absolute value operator. The last inequality follows form the straightforward application of calculus where $Er(X)$ is defined as follows

$$Er(X) \equiv \min\left(\frac{11}{20}, \frac{3}{2\sqrt{\pi}|X|}, \frac{\sqrt{\pi}|X|}{2}\right). \tag{B.21}$$

Combining Eqs. (B.19), (B.20) and (B.21) we get the accuracy result of our approximation:

$$\left|MTF^{e}(u,0) - MTF^{a1}(u,0)\right| \leq \frac{1}{2\sqrt{6ku\alpha}}\left[Er(X_2) + Er(X_1)\right]. \tag{B.22}$$

Considering the construction of this bound it is clear that our approximation performs well only in some ranges of $X_1$ and $X_2$. It turns out that the range of interest in imaging system designs falls within the high accuracy regions of our approximation. To see this let us define $\mathbf{C}$ as the sub-space of interest in imaging system design; namely

$$\mathbf{C} \equiv \{0.2 < u < 1\} \times \{2 < \hat{\alpha} < 10\} \times \left\{0 < \hat{W}_{20} < 8\right\}. \tag{B.23}$$

To see how our approximation accuracy changes with system parameters, let us define $\epsilon$ as the error bound in our approximation; namely

$$\epsilon\left(\hat{\alpha}, \hat{W}_{20}, u\right) \equiv \tag{B.24}$$
$$\frac{1}{2\sqrt{12\pi u\hat{\alpha}}}\left[Er(\sqrt{12\pi u\hat{\alpha}}\, x_m + \sqrt{\frac{4\pi u}{3\hat{\alpha}}}\hat{W}_{20}) + Er(-\sqrt{12\pi u\hat{\alpha}}\, x_m + \sqrt{\frac{4\pi u}{3\hat{\alpha}}}\hat{W}_{20})\right],$$

where $\hat{\alpha} = \alpha/\lambda$ is the normalized cubic phase element and $\hat{W}_{20} = W_{20}/\lambda$ is the normalized defocus. Note that all $\hat{\alpha}$, $\hat{W}_{20}$, and $u$ are dimensionless. Also the arguments of $Er()$ are same as before $X_2$ and $X_1$ but written in terms of the dimensionless variables. Now, we are ready to derive the two important results of this Appendix. In fact, given Eq. (B.24), it is an exercise in calculus to show

that

$$\max_{\mathbf{C}} \left\{ \epsilon \left( \hat{\alpha}, \hat{W}_{20}, u \right) \right\} \leq 0.1, \tag{B.25}$$

$$\frac{1}{\|\mathbf{C}\|} \iiint_{\mathbf{C}} \epsilon \left( \hat{\alpha}, \hat{W}_{20}, u \right) \, \mathrm{d}\hat{\alpha} \, \mathrm{d}\hat{W}_{20} \, \mathrm{d}u \leq 0.03, \tag{B.26}$$

Equation (B.25) states that the accuracy of our approximation is at least 90% and Eq. (B.26) proves that the average accuracy of our approximation is more than 97%. This establishes the high accuracy and practical importance of our approximate analytic expression for MTF. To get an understanding of the behavior of $\epsilon$ we have plotted its value for some ranges of its parameters in Fig. 2-2.

Equation (B.18) as an approximation of the exact MTF traces the average behavior of the exact MTF. One can change this behavior by simply modifying the core approximation equations; namely Eq. (2.3). For example one can see that $|\frac{2}{\pi}\,\mathbf{i}\,\mathrm{Arctan}\left(\sqrt{\frac{\pi}{2}}X\right)$ always remains less than $|-\mathbf{i}\,\mathrm{Erf}[(\mathbf{i}-1)X]|$. Now by choosing the right combination of approximation functions (for $\int_0^{X_2} \exp(\mathbf{i}\,t^2)\mathrm{d}t$ and $\int_0^{X_1} \exp(\mathbf{i}\,t^2)\mathrm{d}t$), one can get the desired approximation of MTF, which always performs as upper bound or lower bound of the exact MTF. This is of particular interest in many design problems when we are looking for the limiting behavior of MTF (such as designs with high safety factor).

# Appendix C

# The $MTF^{a2}$ Approximation

In this Appendix we derive the approximation $MTF^{a2}$ and present some results regarding its accuracy. In particular, we prove the last part of Theorem 2.2.1. To derive the approximation $MTF^{a2}$, we start with Eq. (B.9). In fact, the expression for $I$ in Eq. (B.9) is rather complicated. Although this expression is suitable for calculation, it is not appropriate for analytical optimization. This motivates us to find even a simpler expression for MTF. Thus, we introduce $MTF^{a2}$, the second approximation of MTF by neglecting the complicated parts of $MTF^{a1}$. Note that since $MTF^{a2}$ is always positive in the range of interest, we have dropped the absolute value sign.

$$
\begin{aligned}
MTF^{a2}(u,0) \;=\; & \frac{2}{3\pi^2 ku\alpha}\left[-\mathrm{Arcsin}(u)+\sqrt{\frac{3\pi ku\alpha}{2}}\sqrt{1-u^2}\times\right. \\
& \left(\mathrm{Arctan}\left\{\sqrt{\frac{2\pi ku}{3\alpha}}\,[W_{20}+3\alpha(+1-u)]\right\}\right. \\
& \left.\left.-\mathrm{Arctan}\left\{\sqrt{\frac{2\pi ku}{3\alpha}}\,[W_{20}+3\alpha(-1+u)]\right\}\right)\right].
\end{aligned}
\tag{C.1}
$$

Next, we study the accuracy of $MTF^{a2}$. To study the accuracy of $MTF^{a2}$, we need to compare it with $MTF^{a1}$. This is done numerically in Fig. 2-3. This figure shows the the numerical comparison of the first approximation of MTF and second approximation of MTF. Also, it can rigorously be shown that

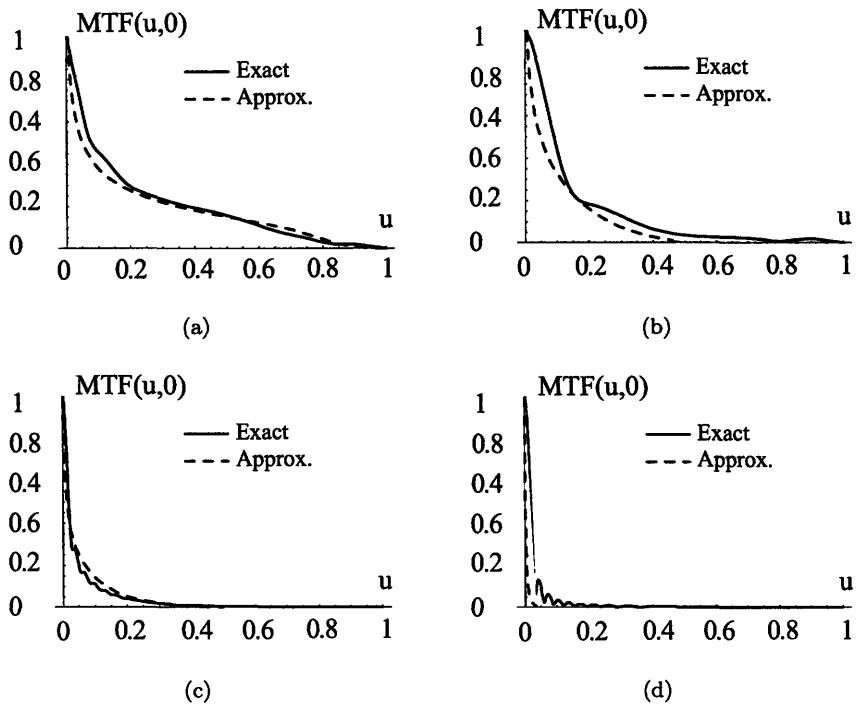



Figure C-1: Plot of exact ($MTF^e$) and approximated $MTF(u,0)$ ($MTF^{a2}$). Note how the exact MTF has many oscillations whereas the approximated MTF does not. Also note that as $\alpha/\lambda$ gets bigger, the accuracy of approximation gets better. (a) $\frac{\alpha}{\lambda}=5, \frac{W_{20}}{\lambda}=1$. (b) $\frac{\alpha}{\lambda}=1, \frac{W_{20}}{\lambda}=1$. (c) $\frac{\alpha}{\lambda}=5, \frac{W_{20}}{\lambda}=5$. (d) $\frac{\alpha}{\lambda}=1, \frac{W_{20}}{\lambda}=5$.

$$\max_{\mathbf{C}} \left| MTF^{a2} - MTF^{a1} \right| \leq 0.1, \tag{C.2}$$

where

$$\mathbf{C} \equiv \{0.2 < u < 1\} \times \{2 < \hat{\alpha} < 10\} \times \left\{0 < \hat{W}_{20} < 8\right\}. \tag{C.3}$$

This serves as the proof of the last part of Theorem 2.2.1. Finally, Fig. C-1 shows the comparison of $MTF^e$ and $MTF^{a2}$ for different values of the imaging system parameters.

# Appendix D

# The Solution of Equation (2.12)

In this Appendix we solve Eq. (2.12). To begin, we rewrite Eq. (2.4) as a function of cubic phase coefficient, as below

$$MTF^{a2}(\hat{\alpha}) = \frac{c_1}{\hat{\alpha}} + \frac{c_2}{\sqrt{\hat{\alpha}}} \times \left[\text{Arctan}(\frac{c_3}{\sqrt{\hat{\alpha}}} + c_4\sqrt{\hat{\alpha}}) - \text{Arctan}(\frac{c_3}{\sqrt{\hat{\alpha}}} - c_4\sqrt{\hat{\alpha}})\right]. \quad \text{(D.1)}$$

where $\hat{\alpha} = \alpha/\lambda$ is the modified cubic phase coefficient and $c_1, \ldots c_4$ are all constants which can be found by comparing Eqs. (D.1) and (2.4), as they are shown in Eqs. (D.2).

$$\begin{aligned} c_1 &= \frac{-\text{Arcsin}(u_{max})}{3\pi^3 u_{max}}, \\ c_2 &= \frac{\sqrt{1-u_{max}^2}}{\pi^2\sqrt{3u_{max}}}, \\ c_3 &= 2\pi\sqrt{\frac{u_{max}}{3}}\frac{W_{20}}{\lambda}, \\ c_4 &= 2\pi\sqrt{3u_{max}}(1-u_{max}), \end{aligned} \quad \text{(D.2)}$$

where $u_{max}$ is defined in Eq. (2.7). Now plugging in the expression for $MTF^{a2}$ from Eq. (D.1) to Eq. (2.12) and performing the differentiation, we have

$$-\frac{c_1}{\hat{\alpha}^2}+\frac{c_2c_4(3{c_3}^2+\hat{\alpha}+{c_4}^2\hat{\alpha}^2)}{\left[\hat{\alpha}+(c_3-c_4\hat{\alpha})^2\right]\left[\hat{\alpha}+(c_3+c_4\hat{\alpha})^2\right]}+ \frac{c_2}{2\hat{\alpha}^{3/2}}\text{Arctan}\left(\frac{-2c_4\hat{\alpha}^{3/2}}{{c_3}^2+\hat{\alpha}-{c_4}^2\hat{\alpha}^2}\right)=0. \tag{D.3}$$

Equation (D.3) is equivalent with Eq. (2.12). Once the numerical value of $c_1,\dots c_4$ are known for a particular problem, Eq. (D.3) can be easily solved using any numerical method. Here we drive an approximate solution for Eq. (D.3). Since $\hat{\alpha}$ is at least more than unity ($\alpha/\lambda>1$), the argument of Arctan is much less than unity and thus $\text{Arctan}(x)\approx x$. Using this, and further simplifying, we get

$$\begin{aligned}&-2c_1({c_3}^2+\hat{\alpha}-{c_4}^2\hat{\alpha}^2)\left[\hat{\alpha}+(c_3-c_4\hat{\alpha})^2\right]\left[\hat{\alpha}+(c_3+c_4\hat{\alpha})^2\right]+\\ &c_2\sqrt{\hat{\alpha}}\Big\{-{c_3}^6\pi+{c_3}^4\hat{\alpha}(-3\pi+4c_4\sqrt{\hat{\alpha}}+3{c_4}^2\pi\hat{\alpha})+\\ &\hat{\alpha}^3(1+{c_4}^2\hat{\alpha})(-\pi-4{c_4}^3\hat{\alpha}^{3/2}+{c_4}^4\pi\hat{\alpha}^2)\\ &+{c_3}^2\hat{\alpha}^2\left[4c_4\sqrt{\hat{\alpha}}+\pi(-3+2{c_4}^2\hat{\alpha}-3{c_4}^4\hat{\alpha}^2)\right]\Big\}=0.\end{aligned} \tag{D.4}$$

The equation above obviously is not analytically solvable, however we can find a correlation for $\hat{\alpha}^*$ considering the range of values of constants $c_1,\dots c_4$ and the range of values of $\hat{\alpha}$. To do so, we first note that $c_1$ and $c_2$ play negligible role in the value of $\hat{\alpha}^*$; i.e. the solution of Eq. (D.4). This is because, $c_1$ is $O(0.01)$ and $c_2$ is $O(0.01)$, whereas $c_3$ is $O(10)$ and $c_4$ is $O(1)$.[1]

Considering this fact, one can perform a term by term order of magnitude analysis on Eq. (D.4) to get a correlation for its solution. By doing so Eq. (D.4) simplifies to

[1]To obtain these orders of magnitude, we have plotted Eqs. (D.2) for $u_{max}\in(0.2,0.8)$ which is the practical interval for $u_{max}$, and observed the range of the constants. We have assumed the average value of $5\lambda$ for defocus coefficient, $W_{20}$.

$$\hat{\alpha}^* \approx \frac{1 + 2c_3c_4 + \sqrt{1 + 4c_3c_4}}{2{c_4}^2}. \tag{D.5}$$

Replacing $c_1 \ldots c_4$ with their corresponding expressions, one can get the solution to Eq. (2.12)

$$\hat{\alpha}^* \approx \frac{1 + 8u_{max}\frac{W_{20}}{\lambda}(1 - u_{max}) + \sqrt{1 + 16u_{max}\frac{W_{20}}{\lambda}(1 - u_{max})}}{24u_{max}(1 - u_{max})^2}. \tag{D.6}$$

# Appendix E

# MTF Approximation Properties

In this Appendix we establish three important properties of our MTF approximation [$MTF^{a2}$ in Eq. (2.4)]. These properties are monotonicity of $MTF^{a2}$ with respect to the normalized spatial frequency $u$ and the defocus ($W_{20}$). We show that as either of these two parameters increase, $MTF^{a2}$ decreases. Also, we show that $MTF^{a2}$ is concave in the domain of $\mathbf{C}$ which is defined by Eq. (B.23) (range of interest in imaging system design).

We first start with defocus. We rewrite Eq. (2.4) as

$$MTF^{a2}(u,0) = \frac{2}{3\pi^2 ku\alpha}\left[-\text{Arcsin}(u) + \sqrt{\frac{3\pi ku\alpha}{2}}\sqrt{1-u^2}\,\Phi\right], \tag{E.1}$$

where

$$\Phi = \text{Arctan}(\chi_1 + \chi_2) - \text{Arctan}(\chi_1 - \chi_2), \tag{E.2}$$

where

$$\begin{aligned} \chi_1 &= \sqrt{\frac{2\pi ku}{3\alpha}}\,(W_{20}), \\ \chi_2 &= \sqrt{6\pi ku\alpha}(1-u). \end{aligned} \tag{E.3}$$

From Eq. (E.1), it is clear that $MTF^{a2}$ is monotonic with respect to $\Phi$ (i.e. $MTF^{a2}$ decreases as $\Phi$ decreases). From Eqs. (E.3), it is clear that $\chi_1$ is proportional to defocus and $\chi_2$ is independent of defocus. So to establish the monotonicity of $MTF^{a2}$ with respect to defocus it remains to show that $\Phi$ is monotonic with respect to $\chi_1$. To do so, we differentiate Eq. (E.2) with respect to $\chi_1$ to get

$$\frac{\mathrm{d}\Phi}{\mathrm{d}\chi_1} = \frac{1}{1+(\chi_1+\chi_2)^2} - \frac{1}{1+(\chi_1-\chi_2)^2}. \tag{E.4}$$

Given the fact that $\chi_1$ and $\chi_2$ are non-negative, it is easy to see that $\mathrm{d}\Phi/\mathrm{d}\chi_1$ is non-positive. This means $\Phi$ is monotonic with respect to $\chi_1$ and it decreases as $\chi_1$ increases. This establishes that $MTF^{a2}$ is monotonic with respect to defocus and it decreases as defocus increases.

Now we move on to the discussion regarding the normalized spatial frequency $u$. To show that $MTF^{a2}$ is monotonic with respect to $u$ and it decreases as $u$ increases, it suffices to show that

$$S = \frac{\mathrm{d}MTF^{a2}(u,0)}{\mathrm{d}u} \leq 0, \tag{E.5}$$

for $u \in (0,1)$, $W_{20} \geq 0$, $d \geq 0$ and $\alpha \geq 0$. We rewrite the Equation for $MTF^{a2}$ (using Eq. (2.4)) as below

$$\begin{aligned} MTF^{a2}(u,0) &= \frac{-2\mathrm{Arcsin}(u)}{3\pi^2 ku\alpha} + \sqrt{\frac{2}{3\pi^3 ku\alpha}}\sqrt{1-u^2}\times \\ &\quad \left\{\mathrm{Arctan}\left[c_1\sqrt{u}+c_2\sqrt{u}(1-u)\right] - \mathrm{Arctan}\left[c_1\sqrt{u}-c_2\sqrt{u}(1-u)\right]\right\}, \end{aligned} \tag{E.6}$$

where

$$\begin{aligned} c_1 &= \sqrt{\frac{2\pi k(W_{20})^2}{3\alpha}}, \\ c_2 &= \sqrt{6\pi k\alpha}. \end{aligned} \tag{E.7}$$

Now, using Eq. (E.6), we can rewrite Eq. (E.5) as

$$\begin{aligned} S(u,c_1,c_2) &= \frac{2}{3\alpha\pi^2 k}\left[\frac{-1}{u\sqrt{1-u^2}}+\frac{\mathrm{Arcsin}(u)}{u^2}\right]+\sqrt{\frac{2}{3\pi^3 k\alpha}}\times \\ &\quad \frac{\mathrm{d}}{\mathrm{d}u}\left(\frac{\sqrt{1-u^2}}{\sqrt{u}}\left\{\mathrm{Arctan}\left[c_1\sqrt{u}+c_2\sqrt{u}(1-u)\right]-\right.\right. \\ &\quad \left.\left.\mathrm{Arctan}\left[c_1\sqrt{u}-c_2\sqrt{u}(1-u)\right]\right\}\right) \end{aligned} \tag{E.8}$$

So the problem is reduced to show $S$ is non-positive for $u \in (0,1)$, $c_1 \geq 0$, $c_2 \geq 0$ and $\alpha \geq 0$. It is straight forward to show (using basic calculus) that the first part of $S$ in brackets, is always non-positive. Thus it suffices to show $S_2$ (which is defined as below) is non-positive.

$$\begin{aligned} S_2(u,c_1,c_2) &= \frac{\mathrm{d}}{\mathrm{d}u}\left(\frac{\sqrt{1-u^2}}{\sqrt{u}}\left\{\mathrm{Arctan}\left[c_1\sqrt{u}+c_2\sqrt{u}(1-u)\right]\right.\right. \\ &\quad \left.\left.-\mathrm{Arctan}\left[c_1\sqrt{u}-c_2\sqrt{u}(1-u)\right]\right\}\right). \end{aligned} \tag{E.9}$$

Since

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}u}\sqrt{1-u^2}=\frac{-u}{\sqrt{1-u^2}}\leq 0, \\ &\frac{1}{\sqrt{u}}\left\{\mathrm{Arctan}\left[c_1\sqrt{u}+c_2\sqrt{u}(1-u)\right]-\mathrm{Arctan}\left[c_1\sqrt{u}-c_2\sqrt{u}(1-u)\right]\right\}\geq 0, \end{aligned} \tag{E.10}$$

it suffices to show $S_3$ (which is defined as below) is non-positive.

$$\begin{aligned} S_3(u,c_1,c_2) &= \frac{\mathrm{d}}{\mathrm{d}u}\left(\frac{1}{\sqrt{u}}\left\{\mathrm{Arctan}\left[c_1\sqrt{u}+c_2\sqrt{u}(1-u)\right]\right.\right. \\ &\quad \left.\left.-\mathrm{Arctan}\left[c_1\sqrt{u}-c_2\sqrt{u}(1-u)\right]\right\}\right). \end{aligned} \tag{E.11}$$

Performing the differentiation we have,

$$\begin{aligned} S_3(u,c_1,c_2) &= \frac{1}{2u^{3/2}}\left\{\frac{2c_2\sqrt{u}\left\{1+u\left[-3-{c_1}^2(1+u)-{c_2}^2(1-u)^2(3u-1)\right]\right\}}{(1+{c_1}^2u)^2-2{c_2}^2u(u-1)^2({c_1}^2u-1)+{c_2}^4u^2(u-1)^4}-\right. \\ &\quad \left.\mathrm{Arctan}\left[c_1\sqrt{u}+c_2\sqrt{u}(1-u)\right]+\mathrm{Arctan}\left[c_1\sqrt{u}-c_2\sqrt{u}(1-u)\right]\right\}, \end{aligned} \tag{E.12}$$

So at this point we only need to show

$$\sup_{u,c_1,c_2} \{S_3(u,c_1,c_2)\} \leq 0, \tag{E.13}$$
$$u \in (0,1)$$
$$c_1 \geq 0$$
$$c_2 \geq 0$$

which is equivalent with

$$\sup_{u,c_1,c_2} \{S_4(u,c_1,c_2)\} \leq 0, \tag{E.14}$$
$$u \in (0,1)$$
$$c_1 \geq 0$$
$$c_2 \geq 0$$

where $S_4$ is defined as

$$\begin{aligned} S_4(u,c_1,c_2) &= \frac{2c_2\sqrt{u}\left\{1+u\left[-3-{c_1}^2(1+u)-{c_2}^2(1-u)^2(3u-1)\right]\right\}}{(1+{c_1}^2u)^2-2{c_2}^2u(u-1)^2({c_1}^2u-1)+{c_2}^4u^2(u-1)^4} - \\ &\quad \text{Arctan}\left[c_1\sqrt{u}+c_2\sqrt{u}(1-u)\right]+\text{Arctan}\left[c_1\sqrt{u}-c_2\sqrt{u}(1-u)\right], \end{aligned} \tag{E.15}$$

We start the maximization by looking at $c_1$ for arbitrary $c_2$ and $u$. From calculus [1] we know that the supremum of $S_4$ with respect to $c_1$ may occur at any critical point, ${c_1}^*$, which: (i) satisfies ${S_4}'({c_1}^*) = 0$, where $'$ represents differentiation with respect to $c_1$, or (ii) ${S_4}'({c_1}^*)$ does not exist, or (iii) ${c_1}^*$ is the boundary of the range of $c_1$ (when the boundary point, ${c_1}^*$, is $\infty$ by supremum occurring at ${c_1}^*$, we mean $\sup S_4(c_1) = \lim_{c_1 \to {c_1}^*} S_4(c_1)$). Considering all these cases we have three

[1] Since $S_4$ is differentiable everywhere in its domain except for finite number of points, we can split its domain to finitely many differentiable open sub-domains. In each sub-domain we can use Theorem 14.3B in [44] to conclude that the supremum in that sub-domain is either (i) $S_4(\vec{x}^*)$, where $\vec{x}^*$ is any of the points at which ${S_4}'(\vec{x})$ vanishes or, (ii) $\lim_{\vec{x}\to\vec{x}_b} S_4(\vec{x})$, where $\vec{x}_b$ is any of the end bounds of the corresponding sub-domain. Since $S_4$ is continuous and the number of non-differentiable points is finite, it remains to evaluate $S_4$ at all of those non-differentiable points and select the maximum among all those values and the supremum of each sub-domain to find the supremum of $S_4$ on its original domain

non-negative and real valued $c_1{}^*$ as below

$$
\begin{aligned}
c_1^{1*} &= 0, \qquad \text{(E.16)}\\
A_1(u,c_2) &= S_4(u,c_1^{1*},c_2) = 2\text{Arctan}\left[c_2\sqrt{u}(u-1)\right] - \frac{2c_2\sqrt{u}(3u-1)}{1+u(1-u)^2{c_2}^2},\\
c_1^{2*} &= \sqrt{(1-u)\left[{c_2}^2(u-1)(2u-1) + u\sqrt{1+4{c_2}^2u^3(1+{c_2}^2u(1-u)^2)}\right]} - 1,\\
A_2(u,c_2) &= S_3(u,c_1^{2*},c_2),\\
c_1^{3*} &= \infty,\\
A_3(u,c_2) &= \lim_{c_1\to\infty} S_4(u,c_1,c_2) = -\pi.
\end{aligned}
$$

Thus, now we need to show that Eqs. (E.17) and (E.18) hold.

$$
\begin{aligned}
&\sup_{u,c_2}\{A_1(u,c_2)\} \le 0. \qquad \text{(E.17)}\\
&u \in (0,1)\\
&c_2 \ge 0
\end{aligned}
$$

$$
\begin{aligned}
&\sup_{u,c_2}\{A_3(u,c_2)\} \le 0. \qquad \text{(E.18)}\\
&u \in (0,1)\\
&c_2 \ge 0
\end{aligned}
$$

We first prove Eq. (E.17). We start the maximization of $A_1(u,c_2)$ by looking at $c_2$ for arbitrary $u$. Following the same method as before the critical points, $c_2{}^*$ are

$$
\begin{aligned}
c_2^{1*} &= 0, \qquad \text{(E.19)}\\
B_1(u) &= A_1(u, c_2^{1*}) = 0,\\
c_2^{2*} &= \frac{1}{(1-u)\sqrt{2u-1}}, \text{ for } u \in (\frac{1}{2}, 1)\\
B_2(u) &= A_1(u, c_2^{2*}) = -2\text{Arctan}\left(\frac{\sqrt{u}}{\sqrt{2u-1}}\right) - \frac{2\sqrt{u(2u-1)}}{1-u},\\
c_2^{3*} &= \infty,\\
B_3(u) &= \lim_{c_2 \to \infty} A_1(u, c_2) = -\pi.
\end{aligned}
$$

It is easy to see that $B_2(u)$ is non-positive for $u \in (\frac{1}{2}, 1)$, and this concludes that Eq. (E.17) is true. Equation (E.18) can be shown to be true following the same line of reasoning.

Thus we have shown that Eq. (E.14) and as a result Eq. (E.5) are true. This means that $MTF^{a2}$ is monotonically decreasing with respect to the normalized spatial frequency $u$.

We skip the proof of the last claim of this Appendix regarding the concavity of $MTF^{a2}$ with domain of $\mathbf{C}$; rather we provide the proof sketch. It is clear that $\mathbf{C}$ is convex by definition in Eq. (B.23). So, it remains to show that the second derivative of $MTF^{a2}$ with respect to $\alpha$ is nonnegative. This last step is a simple exercise in calculus.

# Appendix F

# Derivation of the Expansion for the Point Spread Function

We now derive expressions in Eqs. (3.14) and (3.17) for the PSF. Rather than following the traditional approach of expanding the wavefront phase, $w$, using Zernike polynomials [31], in the first step of analyzing Eq. (3.10), we expand the exponential of the wavefront phase, $\exp(\mathrm{i}\,kw)$, using the set of Zernike polynomials, $V_n^m(\rho,\theta)$

$$V_n^m(\rho,\theta) = R_n^{|m|}(\rho)\mathrm{e}^{\mathrm{i}\,m\theta}. \tag{F.1}$$

Here $n \geq 0$ and $m$ are integers, $n \geq |m|$ and $n-|m|$ is even. Furthermore $R_n^{|m|}(\rho)$ is defined as

$$R_n^{|m|}(\rho) = \sum_{l=0}^{(n-|m|)/2} C_{n,l}^m \rho^{n-2l}, \tag{F.2}$$

where

$$C_{n,l}^m = \frac{(-1)^l (n-l)!}{l![(n+m)/2-l]![(n-m)/2-l]!}. \tag{F.3}$$

Thus we can rewrite the wavefront $\mathrm{e}^{\mathrm{i}\,k\,w(\rho,\theta,r_0,\phi_0)}$ as

$$e^{i\,k\,w(\rho,\theta,r_0,\phi_0)} = \sum_{n=0}^{\infty}\sum_{|m|=0}^{n}\left[\tilde{\delta}_{n-m}A_{nm}R_n^{|m|}(\rho)e^{i\,m\theta}\right], \tag{F.4}$$

where $\tilde{\delta}_i$ is equal to 1 if $i$ is even and 0 otherwise. Considering the properties of Zernike basis functions [31], $A_{nm}$ in Eq. (F.4) is defined as

$$A_{nm} = \frac{n+1}{\pi}\int_0^{2\pi}\int_0^1 e^{i\,k\,w(\rho,\theta,r_0,\phi_0)}\,V_n^m(\rho,\theta)\,\rho\,d\rho\,d\theta. \tag{F.5}$$

Recalling the definition of $\hat{h}$ in Eq. (3.10) and using (F.4), we have

$$\begin{aligned}\hat{h}(x,y;x_0,y_0) &= \frac{1}{2\pi}\int_0^{2\pi}\int_0^1\left\{\sum_{n=0}^{\infty}\sum_{|m|=0}^{n}\left[\tilde{\delta}_{n-m}A_{nm}R_n^{|m|}(\rho)e^{i\,m\theta}\right]\right\} \\ &\quad e^{iR\rho\cos(\theta-\Theta)}\rho\,d\rho\,d\theta.\end{aligned} \tag{F.6}$$

Using the definition of Bessel function [45], we have

$$\int_0^{2\pi} e^{i\,m\theta}e^{iR\rho\cos(\theta-\Theta)}d\theta = 2\pi e^{i\,m\Theta}\,i^{\,m}J_m(R\rho),$$

therefore we can rewrite Eq. (F.6) as

$$\hat{h}(x,y;x_0,y_0) = \int_0^1\left\{\sum_{n=0}^{\infty}\sum_{|m|=0}^{n}\left[\tilde{\delta}_{n-m}A_{nm}R_n^{|m|}(\rho)e^{i\,m\Theta}\,i^{\,m}J_m(R\rho)\right]\right\}\rho\,d\rho. \tag{F.7}$$

Now applying the following property of Zernike polynomials [31]

$$\int_0^1 R_n^{|m|}(\rho)\,J_m(R\rho)\,\rho\,d\rho = (-1)^{\frac{n-m}{2}}\,\frac{J_{n+1}(R)}{R}, \tag{F.8}$$

we can evaluate the integral in Eq. (F.7) as

$$\hat{h}(x,y;x_0,y_0) = \sum_{n=0}^{\infty}\sum_{|m|=0}^{n}\left[\tilde{\delta}_{n-m}A_{nm}e^{i\,m\Theta}\,i^{\,m}(-1)^{\frac{n-m}{2}}\,\frac{J_{n+1}(R)}{R}\right]. \tag{F.9}$$

Now we are concerned with evaluating the Zernike coefficient. This is equivalent to evaluating the integral in Eq. (F.5). Although at first sight it appears that solving Eq. (F.5) and Eq. (3.10) are both equally hard, as it will become clear Eq. (F.5) has several advantages over Eq. (3.10). The most important advantage is that we can evaluate the integral analytically in this new expression without expanding the defocus term, which is shown in Appendix G. This allows us to have the result of arbitrarily large defocused system readily available. Another important property of this method is its fast convergence, which is discussed in Section 3.5 and Appendix H. To solve Eq. (F.5), we first partially expand $e^{i\,k\,w(\rho,\theta,r_0,\phi_0)}$ using Taylor series as follows:

$$
\begin{aligned}
e^{i\,k\,w(\rho,\theta,r_0,\phi_0)} &= e^{i\,k\,\sum_{j=1}^{n_{ab}}\left[f_{L_j,M_j}(r_0)\,(a\,\rho)^{2L_j}\,(a\,\rho\,\cos(\theta-\phi_0))^{M_j}\right]} \\
&= e^{\sum_{j=1}^{n_{ab}}\left[\beta_j\,\rho^{2L_j+M_j}\,\cos^{M_j}(\theta-\phi_0)\right]} \\
&= e^{\sum_{j\in\chi_1}\left[\beta_j\,\rho^{2L_j}\right]}\,e^{\sum_{j\in\chi_2}\left[\beta_j\,\rho^{2L_j+M_j}\,\cos^{M_j}(\theta-\phi_0)\right]} \\
&= e^{\sum_{j\in\chi_1}\left[\beta_j\,\rho^{2L_j}\right]}\prod_{j\in\chi_2} e^{\beta_j\,\rho^{2L_j+M_j}\,\cos^{M_j}(\theta-\phi_0)} \\
&= e^{\sum_{j\in\chi_1}\left[\beta_j\,\rho^{2L_j}\right]}\times \\
&\quad \prod_{j\in\chi_2}\left\{\sum_{N_j=0}^{\infty}\frac{\left[\beta_j\,\rho^{2L_j+M_j}\,\cos^{M_j}(\theta-\phi_0)\right]^{N_j}}{N_j!}\right\},
\end{aligned}
\tag{F.10}
$$

where $N_j$ is the summation variable for each of terms in aberration function which have been Taylor expanded and $\beta_j$ is defined as

$$
\beta_j = i\,k\,f_{L_j,M_j}(r_0)\,a^{2L_j+M_j}. \tag{F.11}
$$

Substituting Eq. (F.10) in Eq. (F.5), we have

$$A_{nm} = \frac{n+1}{\pi}\int_0^{2\pi}\int_0^1 e^{\sum_{j\in\chi_1}\left[\beta_j\,\rho^{2L_j}\right]}\times \prod_{j\in\chi_2}\left\{\sum_{N_j=0}^{\infty}\frac{\left[\beta_j\,\rho^{2L_j+M_j}\,\cos^{M_j}(\theta-\phi_0)\right]^{N_j}}{N_j!}\right\}\times V_n^m(\rho,\theta)\,\rho\,d\rho\,d\theta. \tag{F.12}$$

Integrating over $\theta$ yields

$$A_{nm} = \frac{n+1}{2^{m-1}}\,e^{i\,m\phi_0}\int_0^1 e^{\sum_{j\in\chi_1}\left[\beta_j\,\rho^{2L_j}\right]}R_n^{|m|}(\rho) \left\{\sum_{(\mathbf{N},D)\in\aleph_m} D\prod_{j\in\chi_2}\frac{\left[\beta_j\,\rho^{2L_j+M_j}\right]^{N_j}}{N_j!}\right\}\rho\,d\rho, \tag{F.13}$$

where we have used the orthogonality property of trigonometry functions

$$\begin{aligned}
\forall\, m \le m' = m+2k \quad &: \int_0^{2\pi}\cos(m\theta)\,(\cos(\theta))^{m+2k}\,d\theta = \frac{\pi(m+2k)!}{2^{2k+m-1}k!(m+k)!},\\
\forall\, m < m' = m+2k+1 \quad &: \int_0^{2\pi}\cos(m\theta)\,(\cos(\theta))^{m+2k+1}\,d\theta = 0,\\
\forall\, m > m' \quad &: \int_0^{2\pi}\cos(m\theta)\,(\cos(\theta))^{m'}\,d\theta = 0,\\
\forall\, m, m' \quad &: \int_0^{2\pi}\sin(m\theta)\,(\cos(\theta))^{m'}\,d\theta = 0.
\end{aligned} \tag{F.14}$$

where $m$, $m'$ and $k$ are positive integers. Note that this leaves us with only a few sets of cross terms to deal with as stated in Eq. (F.13) (rather than an infinite number of terms of the Taylor expansion of the aberrations for the term $A_{nm}$). $\mathbf{N} = [N_2, N_3, \ldots N_{n_{ab}}]$ in Eq. (F.13) is a vector containing the values for all $N_j$s. Furthermore, $\aleph_m$, which is a set containing all pairs of vectors in the form $\mathbf{N}$

and scalars $D$, is defined as

$$\aleph_m = \left\{ (\mathbf{N}, D) \mid \sum_{j=2}^{n_{ab}} (M_j N_j) = |m| + 2k,\ D = \frac{(m+2k)!}{2^{2k} k!(m+k)!};\ k, N_j \in \mathcal{N} \right\}. \tag{F.15}$$

Also using the special dependence of Eq. (F.13) on $m$ and by defining $\delta_i$ as one when $m = 0$ and as two otherwise, we can further simplify Eqs. (F.9) and (F.13) as

$$\hat{h}(x, y; x_0, y_0) = \sum_{n=0}^{\infty} \sum_{m=0}^{n} \left\{ \tilde{\delta}_{n-m} \delta_m A_{nm} \cos\left[m(\Theta + \phi_0)\right] \frac{J_{n+1}(R)}{R} \right\}, \tag{F.16}$$

$$A_{nm} = \frac{n+1}{2^{m-1}} \mathrm{i}^n \sum_{(\mathbf{N}, D) \in \aleph_m} D\, \beta_{\mathbf{N}}\, S^m_{n, k_{\mathbf{N}}}(\vec{\beta}), \tag{F.17}$$

and consider $m$ to be a positive integer. Note that $\beta_{\mathbf{N}}$ and $S^m_{n,k_{\mathbf{N}}}(\vec{\beta})$ are defined as

$$S^m_{n,k_{\mathbf{N}}}(\vec{\beta}) = \int_0^1 \prod_{j \in \chi_1} \mathrm{e}^{\beta_j \rho^{2L_j}}\, R^m_n(\rho)\, \rho^{k_{\mathbf{N}}+1}\, d\rho, \tag{F.18}$$

$$\beta_{\mathbf{N}} = \prod_{j \in \chi_2} \frac{(\beta_j)^{N_j}}{N_j!}. \tag{F.19}$$

And we have

$$k_{\mathbf{N}} = \sum_{j \in \chi_2} (2L_j + M_j) N_j. \tag{F.20}$$

Note that considering the definition of $\chi_1$, the defocus term ($\beta_1$) remains unexpanded in the expression of $A_{nm}$ in Eq. (F.13). Furthermore, the only expression which contains defocus term ($\beta_1$) is the $S$ function. Thus, one way to intuitively see that the complexity of our expression of PSF is independent of defocus, is to note that the complexity of the $S$ function is independent of defocus. In Appendix G, we show that the complexity of $S$ is independent of defocus, which allows us to conclude the same about the PSF.

# Appendix G

# Derivation of the $S^m_{n,k_{\mathbf{N}}}(\vec{\beta})$ in Equation (3.19)

Using Eqs. (3.19) and (F.2) and considering the fact that $m > 0$, we have

$$S^m_{n,k_{\mathbf{N}}}(\vec{\beta}) = \int_0^1 \prod_{j\in\chi_1} \mathrm{e}^{\beta_j \rho^{2L_j}} \sum_{l=0}^{(n-m)/2} C^m_{n,l}\rho^{n-2l+k_{\mathbf{N}}+1}\, d\rho. \tag{G.1}$$

Assuming that the summations involved are finite (which means that the number of aberrations under consideration is finite), we have

$$S^m_{n,k_{\mathbf{N}}}(\vec{\beta}) = \sum_{l=0}^{(n-m)/2} C^m_{n,l} \int_0^1 \prod_{j\in\chi_1} \mathrm{e}^{\beta_j \rho^{2L_j}} \rho^{n-2l+k_{\mathbf{N}}+1}\, d\rho. \tag{G.2}$$

Using Eqs. (F.15) and (F.20) we can see that $k_{\mathbf{N}} \geq 0$ and $k_{\mathbf{N}} = m + 2\acute{k}$. On the other hand from the definition of the Zernike polynomials we know that $n + m$ is also an even number. Thus it follows that $n - 2l + k_{\mathbf{N}}$ is an even number. This is an important property of our expansion and a key factor in the derivation that follows. The next step is to replace each term of $\mathrm{e}^{\beta_j \rho^{2L_j}}$ in Eq. (G.2) by its Taylor expansion (except for defocus term ($\beta_1$) which is treated separately)

$$S^m_{n,k_{\mathbf{N}}}(\vec{\beta}) = \sum_{l=0}^{(n-m)/2} C^m_{n,l} \int_0^1 \mathrm{e}^{\beta_1 \rho^2} \prod_{j \in \chi_3} \left[ \sum_{N_j=0}^{\infty} \frac{(\beta_j \rho^{2L_j})^{N_j}}{N_j!} \right] \rho^{n-2l+k_{\mathbf{N}}+1} \, d\rho. \tag{G.3}$$

Thus the solution of the above general form requires the solution of the integral

$$\int_0^1 \mathrm{e}^{\beta_1 \rho^2} \rho^{2\tau+1} \, d\rho, \tag{G.4}$$

where $\tau = \frac{n+k_{\mathbf{N}}}{2} + l + \sum_{j \in \chi_3} L_j N_j$ is an integer number. Note that since $N_j \in \mathcal{N}$, there will be infinitely many fundamental integrals to solve. However to remain in the range of desired accuracy, we only need to use $N_j \in \{0, \ldots, N_j^*\}$, where $N_j^*$ depends on the problem specification and the respective accuracy. Theorem 3.5.1 provides us with an upper bound for $N_j^*$ for a prescribed desired accuracy and problem specifications.

To solve the fundamental integral in Eq. (G.4), we can use the technique of integration by parts $\tau$ times to get

$$\frac{(-\beta_1)^{-(\tau+1)}}{2} \tau! \left[ 1 - \mathrm{e}^{\beta_1} \sum_{k=0}^{\tau} \frac{(-\beta_1)^k}{k!} \right]. \tag{G.5}$$

Note that for virtually any value of $\beta_1$, the number of arithmetic operations necessary in Eq. (G.5) does not increase. This is equivalent with having a method whose complexity is independent of defocus. Using this result, one can find the general formula for $S$. For instance when $\beta_{L,0} = 0$ for $L \geq 2$, we have

$$\begin{aligned} S^m_{n,k_{\mathbf{N}}}(\vec{\beta}) \;=\; & \sum_{l=0}^{(n-m)/2} \left( \frac{C^m_{n,l}}{2} (-\beta_1)^{\frac{-(2+n-2l+k_{\mathbf{N}})}{2}} (\frac{n-2l+k_{\mathbf{N}}}{2})! \right. \\ & \left. \left[ 1 - \mathrm{e}^{\beta_1} \sum_{j=0}^{\frac{n-2l+k_{\mathbf{N}}}{2}} \frac{(-\beta_1)^j}{j!} \right] \right). \end{aligned} \tag{G.6}$$

Here $S^m_{n,k_{\mathbf{N}}}(\vec{\beta})$ is a rational polynomial of order $1+(n+k)/2$ of $\beta_1$. There is also one $\exp(\beta_1)$ factor in the structure of $S^m_{n,k_{\mathbf{N}}}(\vec{\beta})$. Thus we can conclude that in case of primary aberrations $S^m_{n,k_{\mathbf{N}}}(\vec{\beta})$ is

a rational polynomial of order $1+(n+m)/2+{N_{S.A.}}^*+{N_{Coma}}^*$ of $\beta_1$. In general the order of this polynomial is $1+(n+m)/2+\sum_{j\in\chi_4}{N_j}^*$. It follows from this discussion that the number of terms necessary for expanding $S$ function does not change with defocus term ($\beta_1$).

For any given accuracy, there is a value of $|\beta_1|$ for which it may be more efficient to use the method of stationary phase. For instance when $\epsilon = 0.001$, for $|\beta_1| > 700$, it may be more efficient to use the method of stationary phase. In what follows we consider such a case. Note that in practice this is equivalent with the case that the defocus is large enough that the PSF will become almost constant.

Using the fact that $k_N = m + 2\acute{k}$, by changing the variable of $\rho^2$ to $\rho$ and $\beta_1$ to $\mathbf{i}\acute{\beta}$ ($\acute{\beta} \in \Re$), we can rewrite Eq. (G.1) as

$$
\begin{aligned}
S^m_{n,k_{\mathbf{N}}}(\beta_1) &= \sum_{l=0}^{(n-m)/2} C^m_{n,l} \int_0^1 \mathrm{e}^{\mathbf{i}\acute{\beta}\rho} \rho^{\frac{n+m}{2}-l+\acute{k}}\, d\rho \qquad \text{(G.7)}\\
&\approx \sum_{l=0}^{(n-m)/2} C^m_{n,l} \int_0^1 \Big[\cos(\acute{\beta}\rho) + \mathbf{i}\ \rho\ \sin(\acute{\beta}\rho)\Big]\, d\rho\\
&= \sum_{l=0}^{(n-m)/2} C^m_{n,l} \frac{\mathrm{e}^{\mathbf{i}\acute{\beta}}}{\mathbf{i}\acute{\beta}}\\
&= \frac{\mathrm{e}^{\mathbf{i}\acute{\beta}}}{\mathbf{i}\acute{\beta}} \sum_{l=0}^{(n-m)/2} C^m_{n,l}\\
&= \frac{\mathrm{e}^{\mathbf{i}\acute{\beta}}}{\mathbf{i}\acute{\beta}}\\
&= \frac{\mathrm{e}^{\beta_1}}{\beta_1}.
\end{aligned}
$$

In the above derivation, we have assumed that $k_{\mathbf{N}} > 0$; if this is not the case, i.e. $k_{\mathbf{N}} = 0$, then using the same method one can show that

$$
S^m_{n,k_{\mathbf{N}}}(\beta_1) = \frac{\mathrm{e}^{\beta_1} + (-1)^{\frac{n+1}{2}}}{\beta_1}. \qquad \text{(G.8)}
$$

Note that since we do not consider arbitrary large aberrations (except for defocus) we do not

expect large values of $\tau$ in Eq. (G.5) except when many aberrations coexist at the same time. For large values of $\tau$ in Eq. (G.5), one can show that again Eqs. (G.7) and (G.8) may be used to derive $S$. To choose which method to follow (to directly use Eq. (G.5), or to use Eqs. (G.7) and (G.8)) depends on the corresponding value of defocus, $\beta_1$. In fact Eq. (G.7) which is an expression for $S$ function, is a summation of many expressions in the form of Eq. (G.5).

We now discuss the asymptotic behavior of $S$ in the limiting case when $n$ becomes large. When we do not consider arbitrary large aberrations (except for defocus), the asymptotic behavior of Eq. (G.7) for large $n$ (and $m \leq n$) is as follows

$$S^m_{n,k_N}(\vec{\beta}) = \begin{cases} 0 & \text{if } n \neq m, \\ \frac{n!}{2(-\beta_1)^{n+1}} \left[1 - e^{\beta_1} \sum_{k=0}^{n} \frac{(-\beta_1)^k}{k!}\right] & \text{if } n = m, \end{cases} \tag{G.9}$$

Note that even for the case of $n = m$, for very large $n$, $S^m_{n,k_N}(\vec{\beta})$ goes to zero. In fact Eq. (G.9) shows the first-order approximation of the asymptotic behavior of $S^m_{n,k_N}(\vec{\beta})$, when $n$ goes to infinity. This can be seen by noting that large value of $n$ corresponds to large exit window, $R^*$ (see Theorem 3.5.1). Note that when we do not consider arbitrary large aberrations except for defocus, the intensity of light disturbance decreases as $R < R^*$ increases.

# Appendix H

# Complexity Proofs

In this Appendix we will prove Theorem 3.5.1. Before proceeding to the rigorous argument we intuitively explain why the complexity of our representation is independent of defocus. As it is explained in Section 3.5, the two main criteria for assessing the complexity of our representation for a particular accuracy, are the magnitudes of $n^*$ and $N_j^*$ (for $j = 2 \ldots n_{ab}$). In the following discussion, we go over each of these two criteria and show how they are independent of defocus.

We first analyze $N_j^*$ for $j = 2 \ldots n_{ab}$. Considering the last two lines of Eq. (F.10) and Eqs. (G.2) and (G.3). It is clear that $N_j^*$ is the result of expanding the $j^{th}$ aberration exponential. Thus, it should intuitively only depend on $\beta_j$, not on defocus ($\beta_1$).

Regarding $n^*$, consider Eq. (F.4), where the wavefront exponential is first expanded. In Eq. (F.4) both right-hand side and left-hand side are highly oscillatory and they do get more oscillatory as defocus increases. The novel property of the right-hand side expansion is that, once it is substituted in the diffraction integral, higher-order terms will have negligible contribution to the final PSF. Note that this does not mean higher-order terms of right-hand side of Eq. (F.4) have negligible contribution to the right-hand side.

To elaborate this point more, let us call the right-hand side of Eq. (F.4) $\omega_r$ and the left-hand side $\omega_l$. Using Theorem 3.5.1, for accuracy of $\epsilon$, one needs to consider only $n^*$ terms in the right-hand side of Eq. (F.4). This certainly does not mean that right-hand side of Eq. (F.4) will remain within

desired accuracy of the left-hand side as defocus increases. In fact as defocus increases, keeping the number of terms in the right-hand side constant and equal to $n^*$ will make it a poor approximation for the left-hand side. In other words

$$|\omega_r - \omega_l| >> \epsilon. \tag{H.1}$$

Now we substitute each of $\omega_l$ and $\omega_r$ in the diffraction integral (Eq. (3.10)), and call the result $\Omega_l$ and $\Omega_r$. The novel result of this manuscript is that

$$|\Omega_l - \Omega_r| < \epsilon. \tag{H.2}$$

In other words although as defocus increases the difference in Eq. (H.1) increases too, the difference in Eq. (H.2) remains bounded independent of the value of defocus. This property could be intuitively explained by investigating the property of higher order Airy functions. In particular, as the order of these functions increases their contributions to the summation to which they belong decreases. This observation leads to the claim in Eq. (H.2).

We begin the proof of Theorem 3.5.1 by presenting four lemmas.

**Lemma H.0.1.** *For all $A_{nm}$ defined by Eq. (3.17), we have the following bound*

$$|A_{nm}| \leq \sqrt{n+1}. \tag{H.3}$$

**Proof:** Multiplying both sides of Eq. (F.4) by their complex conjugate and then performing an integration over the unit circle yields

$$2\pi {A_{00}}^2 + \sum_{n=1}^{\infty} \sum_{|m|=1}^{n} \left[ \tilde{\delta}^2_{n-m} A^2_{nm} \frac{\pi}{n+1} \right] = 2\pi \tag{H.4}$$

where we have used the orthogonality property of the Zernike basis functions over the unit circle. Since $A_{n,m}$ and $A_{n,-m}$ are the same, we have

$$2\pi {A_{00}}^2 + \sum_{n=1}^{\infty} \sum_{m=1}^{n} \left[ 2\tilde{\delta}_{n-m} A_{nm}^2 \frac{\pi}{n+1} \right] = 2\pi. \tag{H.5}$$

Simplifying Eq. (H.5), we have

$${A_{00}}^2 + \sum_{n=1}^{\infty} \sum_{m=1}^{n} \left[ \tilde{\delta}_{n-m} A_{nm}^2 \frac{1}{n+1} \right] = 1. \tag{H.6}$$

Lemma H.0.1 follows from Eq. (H.6) immediately:

$$|A_{nm}| \leq \sqrt{n+1}. \tag{H.7}$$

□

Let $T_n^f(x)$ be the $n^{th}$ term in the Taylor expansion of $f(x)$ around $x_0 = 0$, then we have the following Lemma about the accuracy of the Taylor expansion.

**Lemma H.0.2.** *Let*

$$f(x) = \mathrm{e}^{b\,x^m},$$

*where $b$ is an imaginary number and $x \in [0,1]$. If*

$$p^* = \max\left(4,\, 2\mathrm{e}\,|b| + 1,\, \log_2 \frac{\mathrm{e}}{(2\mathrm{e}-1)\sqrt{2\pi\epsilon}}\right).$$

*Then we have*

$$\left| f(x) - \sum_{n=0}^{mp^*} T_n^f(x) \right| \leq \epsilon \left| f(x) \right| .$$

**Proof:** Using Taylor Theorem we have

$$\begin{aligned} \left| f(x) - \sum_{n=0}^{mp^*} T_n^f(x) \right| &= \left| \sum_{n=mp^*+1}^{\infty} T_n^f(x) \right| \\ &= \left| \sum_{n=mp^*+1}^{\infty} \frac{f^n(0)}{n!} x^n \right| . \end{aligned} \tag{H.8}$$

Using the definition of $f$, we have

$$f^n(0) = \begin{cases} 0 & \text{if } n \neq mp \ , \\ n! \frac{b^p}{p!} & \text{if } n = mp. \end{cases} \tag{H.9}$$

where $p$ is a positive integer. Substituting this in Eq. (H.8) and changing the index of summation from $n$ to $p$, we have

$$
\begin{aligned}
\left| f(x) - \sum_{n=0}^{mp^*} T_n^f(x) \right| &= \left| \sum_{p=p^*}^{\infty} \frac{b^p}{p!} x^{mp} \right| \qquad \text{(H.10)} \\
&\leq \left| \sum_{p=p^*}^{\infty} \frac{b^p}{p!} \right| \\
&\leq \sum_{p=p^*}^{\infty} \frac{|b|^p}{p!} \\
&\leq \sum_{p=p^*}^{\infty} \frac{|b|^p}{p^*!\, p^{*p-p^*}} \\
&= \frac{|b|^{p^*}}{p^*!} \sum_{p=p^*}^{\infty} \left( \frac{|b|}{p^*} \right)^p \\
&\leq \frac{|b|^{p^*}}{p^*!} \sum_{p=p^*}^{\infty} \left( \frac{1}{2\mathrm{e}} \right)^p \\
&= \frac{|b|^{p^*}}{p^*!} \frac{2\mathrm{e}}{2\mathrm{e}-1} \\
&\leq \frac{|b|^{p^*}}{\sqrt{2\pi} p^{*p^*+0.5} \mathrm{e}^{-p^*}} \frac{2\mathrm{e}}{2\mathrm{e}-1} \\
&= \left( \frac{|b|\,\mathrm{e}}{p^*} \right)^{p^*} \frac{2\mathrm{e}}{(2\mathrm{e}-1)\sqrt{2\pi p^*}} \\
&\leq \left( \frac{1}{2} \right)^{p^*} \frac{2\mathrm{e}}{(2\mathrm{e}-1)\sqrt{2\pi p^*}} \\
&\leq \left( \frac{1}{2} \right)^{p^*} \frac{\mathrm{e}}{(2\mathrm{e}-1)\sqrt{2\pi}} \\
&\leq \epsilon \\
&\leq |f(x)|\, \epsilon.
\end{aligned}
$$

The first inequality follows from $x \in [0,1]$. The third inequality follows from $p! \geq p^*! p^{*p-p^*}$. The fourth inequality follows from $p^* \geq 2\mathrm{e}\,|b| + 1$. In the fifth inequality, we have applied the following lower bound to $p^*!$, based on Stirling's approximation:

$$\sqrt{2\pi} p^{*p^*+0.5} \exp(-p^*) \leq p^*!.$$

The sixth inequality follows from $p^* \geq 2e\,|b| + 1$. The seventh inequality follows from $p^* \geq 4$. In the eighth inequality we have used $p^* \geq \log_2 \frac{\mathrm{e}}{(2\mathrm{e}-1)\sqrt{2\pi\epsilon}}$. In the ninth inequality we have used

$|\exp(\mathrm{i}\,x)| = 1$ for all $x \in \Re$.

□

**Lemma H.0.3.** *For all $R \in \Re$ we have*

$$\sum_{n=0}^{\infty}(n+1)^{3/2}\,|J_{n+1}(R)| \leq \sqrt{\frac{3}{\pi}}\,\mathrm{e}^2\left(1+R^{4/3}\right)R.$$

**Proof:** The first kind $\mathrm{n}^{\mathrm{th}}$ order Bessel functions have two classical bounds [46, 47, 48]

$$\begin{aligned} |J_n(R)| &\leq \sqrt{\frac{2}{\pi}}\,\frac{1}{R^{1/3}} \\ |J_n(R)| &\leq \frac{(R/2)^n}{n!}. \end{aligned} \tag{H.11}$$

where in the second one, $\sqrt{\frac{2}{\pi}} > 0.7857\ldots$, the constant derived by Landau [48]. Since these two bounds are always true, we can define $f_n(R)$, a special upper bound for the first kind $n^{th}$ order Bessel function, as

$$|J_n(R)| \leq f_n(R) = \begin{cases} \frac{(R/2)^n}{n!} & \text{for } 0 \leq R < \frac{n}{\mathrm{e}}, \\ \\ \sqrt{\frac{2}{\pi}}\frac{1}{R^{1/3}} & \text{for } R \geq \frac{n}{\mathrm{e}}, \end{cases} \tag{H.12}$$

Using this equation, we have

$$\begin{aligned} \sum_{n=0}^{\infty}(n+1)^{3/2}\,|J_{n+1}(R)| &\leq \sum_{n=1}^{\lfloor \mathrm{e}R \rfloor} n^{3/2}\sqrt{\frac{2}{\pi}}\frac{1}{R^{1/3}} + \sum_{n=\lfloor \mathrm{e}R \rfloor+1}^{\infty} n^{3/2}\frac{(R/2)^n}{n!} \\ &= I_1 + I_2. \end{aligned} \tag{H.13}$$

Now we consider each term in the right-hand-side of Eq. (H.13) separately. For the first term we note that for $0 \leq R < \frac{1}{\mathrm{e}}$, we have

$$I_1 = 0. \tag{H.14}$$

Now, we assume $R \geq \frac{1}{\mathrm{e}}$ and we have

$$\begin{aligned} I_1 &= \sqrt{\frac{2}{\pi}} \frac{1}{R^{1/3}} \sum_{n=1}^{\lfloor \mathrm{e}R \rfloor} n^{3/2} \\ &\leq \sqrt{\frac{2}{\pi}} \frac{1}{R^{1/3}} \int_1^{\mathrm{e}R+1} x^{3/2} dx \\ &= \frac{2\sqrt{2}}{5\sqrt{\pi}} \frac{(\mathrm{e}R+1)^{5/2}-1}{R^{1/3}} \\ &\leq \frac{2\sqrt{2}}{5\sqrt{\pi}} \frac{1}{R^{1/3}} \mathrm{e}\sqrt{\mathrm{e}}(R)^{4/3} \left[1+(\mathrm{e}R+2)^{4/3}\right] \\ &= \frac{2\mathrm{e}\sqrt{2\mathrm{e}}}{5\sqrt{\pi}} R \left[1+(\mathrm{e}R+2)^{4/3}\right]. \end{aligned} \tag{H.15}$$

The first inequality follows from the definition of integration. The second inequality follows from $(\mathrm{e}R+1)^{5/2}-1 \leq \mathrm{e}\sqrt{\mathrm{e}}R^{4/3}\left[1+(\mathrm{e}R+2)^{4/3}\right]$ for $R \geq \frac{1}{\mathrm{e}}$. Putting together Eqs. (H.14) and (H.15), we have

$$I_1 \leq \frac{2\mathrm{e}\sqrt{2\mathrm{e}}}{5\sqrt{\pi}} R \left[1+(\mathrm{e}R+2)^{4/3}\right]. \tag{H.16}$$

for all $R \geq 0$. As for $I_2$ we first consider $R \geq 1$ as

$$
\begin{aligned}
I_2 &= \sum_{n=\lfloor eR\rfloor+1}^{\infty} n^{3/2}\frac{(R/2)^n}{n!} \qquad \text{(H.17)}\\
&\le \sum_{n=\lfloor eR\rfloor+1}^{\infty} \frac{(R/2)^n}{(n-2)!}\\
&= \frac{R^2}{4}\sum_{n=\lfloor eR\rfloor-1}^{\infty} \frac{(R/2)^n}{n!}\\
&\le \frac{R^2}{4}\sum_{n=\lfloor eR\rfloor-1}^{\infty} \frac{(R/2)^n}{(\lfloor eR\rfloor-1)!(\lfloor eR\rfloor-1)^{n-(\lfloor eR\rfloor-1)}}\\
&= \frac{R^2(R/2)^{(\lfloor eR\rfloor-1)}}{4(\lfloor eR\rfloor-1)!}\sum_{n=0}^{\infty}\left[\frac{R}{2(\lfloor eR\rfloor-1)}\right]^n\\
&\le \frac{R^2(R/2)^{(\lfloor eR\rfloor-1)}}{4(\lfloor eR\rfloor-1)!}\sum_{n=0}^{\infty}\left(\frac{1}{e-1}\right)^n\\
&= \frac{(e-1)R^2(R/2)^{(\lfloor eR\rfloor-1)}}{4(e-2)(\lfloor eR\rfloor-1)!}\\
&\le \frac{(e-1)R^2\left[\frac{R}{2(\lfloor eR\rfloor-1)}\right]^{\lfloor eR\rfloor-1}}{4(e-2)\sqrt{2\pi(\lfloor eR\rfloor-1)}}\\
&\le \frac{(e-1)R^{3/2}\left(\frac{1}{e-1}\right)^{\lfloor eR\rfloor-1}}{4(e-2)\sqrt{2\pi}}\sqrt{\frac{2}{e-1}}\\
&= \frac{\sqrt{e-1}R^{3/2}\left(\frac{1}{e-1}\right)^{\lfloor eR\rfloor-1}}{4(e-2)\sqrt{\pi}}\\
&\le \frac{\sqrt{e-1}R^{3/2}\left(\frac{1}{e-1}\right)^{eR-1}}{4(e-2)\sqrt{\pi}}\\
&\le \frac{(e-1)\sqrt{e-1}R}{4(e-2)\sqrt{\pi}}\\
&\le \frac{e-1}{2\sqrt{\pi}(e-2)}R.
\end{aligned}
$$

The first inequality follows from $n^{3/2} \le n(n-1)$ for $n \ge 3$. The second inequality follows from $n! \ge (\lfloor eR\rfloor-1)!(\lfloor eR\rfloor-1)^{n-(\lfloor eR\rfloor-1)}$. In the third and fifth inequality we have used $\frac{R}{2(\lfloor eR\rfloor-1)} \le \frac{1}{e-1}$ for $R \ge 1$. In the fourth inequality, we have applied the following lower bound to $(\lfloor eR\rfloor-1)!$, based on Stirling's approximation:

$$
\sqrt{2\pi}(\lfloor eR\rfloor-1)^{\lfloor eR\rfloor-1+0.5}\exp\left[-(\lfloor eR\rfloor-1)\right] \le (\lfloor eR\rfloor-1)!
$$

In the fifth inequality we have also used $\sqrt{\frac{R}{2(\lfloor eR \rfloor - 1)}} \leq \sqrt{\frac{2}{e-1}}$ for $R \geq 1$. In the sixth inequality we have used $x \geq \lfloor x \rfloor$ for all $x$. The seventh inequality follows from $\sqrt{R} \geq (e-1)^{eR}$ for $R \geq 1$.

In a similar way, it could be shown that the bound for $I_2$ in Eq. (H.17) holds for $0 \leq R < 1$ too. Now by combining Eqs (H.13), (H.16) and (H.17) together, we have

$$\begin{aligned}
\sum_{n=0}^{\infty}(n+1)^{3/2}\,|J_{n+1}(R)| &\leq \frac{2e\sqrt{2e}}{5\sqrt{\pi}}R\left[1+(eR+2)^{4/3}\right]+\frac{e-1}{2\sqrt{\pi}(e-2)}R \qquad \text{(H.18)}\\
&= \frac{R}{\sqrt{\pi}}\left\{\frac{2e\sqrt{2e}}{5}\left[1+(eR+2)^{4/3}\right]+\frac{e-1}{2(e-2)}\right\}\\
&\leq \frac{R}{\sqrt{\pi}}\left[\sqrt{3}e^2\left(1+R^{4/3}\right)\right]\\
&= \sqrt{\frac{3}{\pi}}\,e^2\left(1+R^{4/3}\right)R.
\end{aligned}$$

□

Before presenting Lemma H.0.4 and proof of Theorem 3.5.1 we define some relevant parameters. Let $A_{nm}$ be the exact values of coefficients in Eq. (3.14) from Eq. (3.17) and let $A^*_{nm}$ be the approximated value of these coefficients from Eq. (3.35) as stated in Eqs. (H.19) and (H.20) respectively.

$$A_{nm} = \frac{n+1}{\pi}\int_0^{2\pi}\int_0^1 e^{\beta_1\rho^2}\prod_{j=2}^{n_{ab}} f_{L_j,M_j}(\rho,\theta)\,V_n^m(\rho,\theta)\,\rho\,d\rho\,d\theta. \qquad \text{(H.19)}$$

$$A^*_{nm} = \frac{n+1}{\pi}\int_0^{2\pi}\int_0^1 e^{\beta_1\rho^2}\prod_{j=2}^{n_{ab}} T_{L_j,M_j}(\rho,\theta)\,V_n^m(\rho,\theta)\,\rho\,d\rho\,d\theta. \qquad \text{(H.20)}$$

Also let $f_{L_j,M_j}(\rho,\theta)$, $T_{L_j,M_j}(\rho,\theta)$ and $\epsilon_{L_j,M_j}(\rho,\theta)$ be short form expression of the exponential factor of the aberration $(L_j, M_j)$, $\exp\left[\beta_j\rho^{2L_j+M_j}\cos^{M_j}(\theta-\phi_0)\right]$, the Taylor expansion of this exponential expression and the error of this expansion corresponding to the first $N_j{}^*$ terms of the expansion respectively. Also let

$$\hat{h}(x,y;x_0,y_0) = \sum_{n=0}^{\infty}\sum_{m=1}^{n}\tilde{\delta}_{n-m}\delta_m A_{nm}\cos[(\Theta+\phi_0)]\frac{J_{n+1}(R)}{R}, \tag{H.21}$$

$$\hat{h}_{n^*}(x,y;x_0,y_0) = \sum_{n=0}^{n^*}\sum_{m=1}^{n}\tilde{\delta}_{n-m}\delta_m A^*_{nm}\cos[(\Theta+\phi_0)]\frac{J_{n+1}(R)}{R},$$

be the normalized exact and approximated PSF respectively. There, $\delta_m$ is one if $m = 0$ and it is two otherwise and $\tilde{\delta}_i$ is one when $i$ is even and it is zero otherwise. We recall that $\hat{h}(x,y;x_0,y_0)$ in Eq. (H.21) is the exact PSF and $\hat{h}_{n^*}(x,y;x_0,y_0)$ is the approximate PSF whose accuracy is under investigation.

Let $R^*$ be the radius of the region of interest (exit window) and let $n_{ab}$ be the number of aberrations present in an optical system. We will prove that to have an arbitrarily accurate result in this region we only need a minimum necessary index of summation, $n^*$, in Eq. (H.21) and a minimum necessary number of terms of Taylor expansion for each aberration, ${N_j}^*$, in Eq. (H.20). Note that both summation indices are independent of the value of defocus.

**Lemma H.0.4.** *Suppose that we have finitely many aberrations ($n_{ab}$) in an optical system. If*

$$f_{L_j,M_j}(\rho,\theta) - T_{L_j,M_j}(\rho,\theta) = -\epsilon_{L_j,M_j}(\rho,\theta)f_{L_j,M_j}(\rho,\theta), \tag{H.22}$$

*and*

$${N_j}^* = \max\left(4,\, 2\mathrm{e}\,|\beta_j| + 1,\, \log_2\frac{\mathrm{e}}{(2\mathrm{e}-1)\sqrt{2\pi\epsilon}}\right). \tag{H.23}$$

*for all $j = 2\ldots n_{ab}$ in the system, then we have*

$$\left|\epsilon_{L_j,M_j}\right| \leq \epsilon.$$

*for all* $j = 2 \ldots n_{ab}$.

**Proof:** Referring to Eq. (H.22), we set $p^* = N_j^*$, $b = \beta_j \cos^{M_j}(\theta - \phi_0)$, $x = \rho$ and $m = 2L_j + M_j$. We can rewrite Eq. (H.23) to get

$$p^* = \max\left(4,\, 2\mathrm{e}\,|b| + 1,\, \log_2 \frac{\mathrm{e}}{(2\mathrm{e}-1)\sqrt{2\pi\epsilon}}\right). \tag{H.24}$$

Now using Lemma H.0.2, we have

$$\left|f_{L_j,M_j}(\rho,\theta) - T_{L_j,M_j}(\rho,\theta)\right| \leq \epsilon \left|f_{L_j,M_j}(\rho,\theta)\right|. \tag{H.25}$$

Substituting the left-hand-side from Eq. (H.22), we have

$$\left|\epsilon_{L_j,M_j}(\rho,\theta) f_{L_j,M_j}(\rho,\theta)\right| \leq \epsilon \left|f_{L_j,M_j}(\rho,\theta)\right|. \tag{H.26}$$

Simplifying Eq. (H.26) yields

$$\left|\epsilon_{L_j,M_j}\right| \leq \epsilon. \tag{H.27}$$

□

**Proof of Theorem 3.5.1:** We first define the error of Taylor expansion, $\epsilon_{L,M}(\rho,\theta)$, as in Lemma H.0.4

$$\epsilon_{L_j,M_j}(\rho,\theta) = \frac{T_{L_j,M_j}(\rho,\theta) - f_{L_j,M_j}(\rho,\theta)}{f_{L_j,M_j}(\rho,\theta)}. \tag{H.28}$$

Using the definition of $A_{nm}$ and $A^*_{nm}$ we have (from here on for the sake of simplicity we will not write the dependence of $f_{L,M}$, $T_{L,M}$, $\epsilon_{L,M}$ and $V$ on $\rho$ and $\theta$ explicitly)

$$\begin{aligned} A_{nm} - A^*_{nm} &= \frac{n+1}{\pi} \int_0^{2\pi} \int_0^1 \left( \prod_{j=2}^{n_{ab}} f_{L_j,M_j} - \prod_{j=2}^{n_{ab}} T_{L_j,M_j} \right) \mathrm{e}^{\beta_1 \rho^2} V_n^m \, \rho \, d\rho \, d\theta \\ &= \frac{n+1}{\pi} \int_0^{2\pi} \int_0^1 \left[ \prod_{j=2}^{n_{ab}} f_{L_j,M_j} - \prod_{j=2}^{n_{ab}} f_{L_j,M_j}(1+\epsilon_{L_j,M_j}) \right] \mathrm{e}^{\beta_1 \rho^2} V_n^m \, \rho \, d\rho \, d\theta \\ &= \frac{n+1}{\pi} \int_0^{2\pi} \int_0^1 \left[ \left( \prod_{j=2}^{n_{ab}} f_{L_j,M_j} \right) g \right] \mathrm{e}^{\beta_1 \rho^2} V_n^m \, \rho \, d\rho \, d\theta \end{aligned} \tag{H.29}$$

In the second equality we have used Eq. (H.28). The third equality follows from the definition of $g$ in Eq. (H.30).

$$\begin{aligned} g &= 1 - \prod_{j=2}^{n_{ab}} (1+\epsilon_{L_j,M_j}) \\ &= \sum_{j=2}^{n_{ab}} \epsilon_{L_j,M_j} + \sum_{j,k=2}^{n_{ab}} \epsilon_{L_j,M_j} \epsilon_{L_k,M_k} + \cdots + \prod_{j=2}^{n_{ab}} \epsilon_{L_j,M_j}. \end{aligned} \tag{H.30}$$

Taking absolute values, we have

$$\begin{aligned} |g| &\leq \left| \sum_{j=2}^{n_{ab}} \epsilon_{L_j,M_j} \right| + \left| \sum_{j,k=2}^{n_{ab}} \epsilon_{L_j,M_j} \epsilon_{L_k,M_k} \right| + \cdots + \left| \prod_{j=2}^{n_{ab}} \epsilon_{L_j,M_j} \right| \\ &\leq \binom{n_{ab}-1}{1} \epsilon' + \binom{n_{ab}-1}{2} \epsilon'^2 + \cdots + \binom{n_{ab}-1}{n_{ab}-1} \epsilon'^{n_{ab}-1} \\ &\leq \binom{n_{ab}}{1} \epsilon' + \binom{n_{ab}}{2} \epsilon'^2 + \cdots + \binom{n_{ab}}{n_{ab}} \epsilon'^{n_{ab}} \\ &\leq n_{ab} \epsilon' \left( 1 + \frac{(1/2)^2}{2!} + \cdots + \frac{(1/2)^{n_{ab}}}{n_{ab}!} \right) \\ &\leq n_{ab} \epsilon' \sum_{j=0}^{\infty} \frac{(1/2)^j}{j!} \\ &= n_{ab} \epsilon' \sqrt{e} \end{aligned} \tag{H.31}$$

The second inequality follows from applying the Lemma H.0.4 by setting

$$\epsilon^{'} = \frac{\sqrt{\pi}}{2\sqrt{3\mathrm{e}}\,\mathrm{e}^2\, n_{ab}(1+R^{*4/3})}\,\epsilon, \tag{H.32}$$

as the desired accuracy. Note that the value of $N_j^*$ required for this accuracy is precisely what is stated in the expression of Theorem 3.5.1. The fourth inequality follows from $n_{ab}\epsilon^{'} \leq \frac{1}{2}$, which in turn follows from Eq. (H.32) by noting that $\epsilon \leq 1$ and $R^* \geq 0$. Now, taking absolute values in Eq. (H.29), we have

$$\begin{aligned} |A_{nm} - A_{nm}^*| &\leq \frac{n+1}{\pi}\int_0^{2\pi}\int_0^1 |g|\,|R_n^m(\rho)|\;\rho\,d\rho\,d\theta \\ &\leq \frac{n_{ab}\epsilon^{'}\sqrt{\mathrm{e}}(n+1)}{\pi}\int_0^{2\pi}\int_0^1 |R_n^m(\rho)|\;\rho\,d\rho\,d\theta \\ &= 2n_{ab}\epsilon^{'}\sqrt{\mathrm{e}}(n+1)\int_0^1 |R_n^m(\rho)|\;\rho\,d\rho\,d\theta \\ &\leq 2n_{ab}\epsilon^{'}\sqrt{\mathrm{e}}(n+1)\frac{1}{2\sqrt{n+1}} \\ &= n_{ab}\epsilon^{'}\sqrt{\mathrm{e}}\sqrt{n+1}. \end{aligned} \tag{H.33}$$

The second inequality follows from Eq. (H.31). The third inequality follows from applying Cauchy-Schwarz inequality to $\int_0^1 [R_n^m(\rho)]^2\rho\,d\rho = \frac{1}{2(n+1)}$ [31]. Now using the definition of the normalized exact and approximated PSF in Eq. (H.21), we have

$$\begin{aligned} \hat{h}(x,y;x_0,y_0) - \hat{h}_{n^*}(x,y;x_0,y_0) = & \\ \sum_{n=0}^{n^*}\sum_{m=0}^{n} \delta_m\tilde{\delta}_{n-m}\,(A_{nm} - A_{nm}^*)\cos\left[m(\Theta+\phi_0)\right]\frac{J_{n+1}(R)}{R} & \\ + \sum_{n=n^*+1}^{\infty}\sum_{m=0}^{n} \delta_m\tilde{\delta}_{n-m}A_{nm}\cos\left[m(\Theta+\phi_0)\right]\frac{J_{n+1}(R)}{R}. & \end{aligned} \tag{H.34}$$

Taking absolute values, we have

$$
\begin{aligned}
&\left|\hat{h}(x,y;x_0,y_0) - \hat{h}_{n^*}(x,y;x_0,y_0)\right| \leq \qquad \text{(H.35)}\\
&\left|\sum_{n=0}^{n^*}\sum_{m=0}^{n} \delta_m \tilde{\delta}_{n-m} \left(A_{nm} - A^*_{nm}\right) \cos\left[m(\Theta+\phi_0)\right] \frac{J_{n+1}(R)}{R}\right| \\
&+\left|\sum_{n=n^*+1}^{\infty}\sum_{m=0}^{n} \delta_m \tilde{\delta}_{n-m} A_{nm} \cos\left[m(\Theta+\phi_0)\right] \frac{J_{n+1}(R)}{R}\right| \\
=\;& I_1 + I_2.
\end{aligned}
$$

Now we analyze each part of the right-hand-side of Eq. (H.35). After rearranging, the first term, $I_1$, simplifies to

$$
\begin{aligned}
I_1 &\leq \sum_{n=0}^{n^*}\sum_{m=0}^{n} \delta_m \tilde{\delta}_{n-m} \left|A_{nm} - A^*_{nm}\right| \left|\frac{J_{n+1}(R)}{R}\right| \qquad \text{(H.36)}\\
&\leq \sum_{n=0}^{n^*}\sum_{m=0}^{n} \delta_m \tilde{\delta}_{n-m} n_{ab}\epsilon' \sqrt{\mathrm{e}}\sqrt{n+1} \left|\frac{J_{n+1}(R)}{R}\right| \\
&= n_{ab}\epsilon' \sqrt{\mathrm{e}} \sum_{n=0}^{n^*} (n+1)^{3/2} \left|\frac{J_{n+1}(R)}{R}\right| \\
&= \frac{n_{ab}\epsilon' \sqrt{\mathrm{e}}}{R} \sum_{n=0}^{n^*} (n+1)^{3/2} \left|J_{n+1}(R)\right| \\
&\leq \frac{n_{ab}\epsilon' \sqrt{\mathrm{e}}}{R} \sqrt{\frac{3}{\pi}}\, \mathrm{e}^2 \left(1 + R^{4/3}\right) R \\
&= \frac{n_{ab}\epsilon' \mathrm{e}^2 \sqrt{3\mathrm{e}}}{\sqrt{\pi}} \left(1 + R^{4/3}\right) \\
&\leq \frac{n_{ab}\epsilon' \mathrm{e}^2 \sqrt{3\mathrm{e}}}{\sqrt{\pi}} \left(1 + R^{*4/3}\right) \\
&= \frac{\epsilon}{2}.
\end{aligned}
$$

The second inequality follows from Eq. (H.33). The third inequality follows from Lemma H.0.3. The fourth inequality follows from $R \leq R^*$. The last equality follows from Eq. (H.32).

By considering the fact that $n - m$ is always even in Zernike polynomials and noting that $\tilde{\delta}_{n-m}$ is equal to zero when $n - m$ is odd, we can further simplify the summation over $m$ in $I_2$ in Eq. (H.35) to get

$$I_2 \leq \sum_{n=n^*+1}^{\infty} \left[ \sqrt{n+1} \left| \frac{J_{n+1}(R)}{R} \right| \left( \tilde{\delta}_n + 2 \sum_{m=1}^{n} \tilde{\delta}_{n-m} |\cos[m(\Theta + \phi_0)]| \right) \right] . \tag{H.37}$$

We can further simplify Eq. (H.37) to get

$$I_2 \leq \sum_{n=n^*+1}^{\infty} \left[ (n+1)^{\frac{3}{2}} \left| \frac{J_{n+1}(R)}{R} \right| \right] . \tag{H.38}$$

Now we can use Eq. (3.18) in Lemma H.0.3 to get

$$I_2 \leq \sum_{n=n^*+1}^{\infty} \left[ (n+1)^{\frac{3}{2}} \frac{1}{2n!} \left( \frac{R}{2} \right)^{n-1} \right] . \tag{H.39}$$

Since for $n \geq 5$ we have $n(n-1) > (n+1)^{\frac{3}{2}}$, we can rewrite Eq. (H.39) as

$$I_2 \leq \frac{R}{4} \sum_{n=n^*-1}^{\infty} \left[ \frac{1}{n!} \left( \frac{R}{2} \right)^{n} \right] . \tag{H.40}$$

From Eq. (H.40), it is clear that the right-hand-side reaches its maximum, when $R = R^*$. Thus

$$
\begin{aligned}
I_2 &\leq \frac{R^*}{4} \sum_{n=n^*-1}^{\infty} \left[ \frac{1}{n!} \left( \frac{R^*}{2} \right)^n \right] \qquad \text{(H.41)} \\
&\leq \frac{R^*}{4} \sum_{n=n^*-1}^{\infty} \frac{\left( \frac{R^*}{2} \right)^n}{(n^*-1)!(n^*-1)^{n-(n^*-1)}} \\
&= \frac{R^*}{4} \frac{\left( \frac{R^*}{2} \right)^{n^*-1}}{(n^*-1)!} \sum_{n=0}^{\infty} \left( \frac{R^*}{2(n^*-1)} \right)^n \\
&\leq \frac{R^*}{4} \frac{\left( \frac{R^*}{2} \right)^{n^*-1}}{(n^*-1)!} \sum_{n=0}^{\infty} \left( \frac{1}{2\mathrm{e}} \right)^n \\
&= \frac{\mathrm{e}R^*}{2(2\mathrm{e}-1)} \frac{\left( \frac{R^*}{2} \right)^{n^*-1}}{(n^*-1)!} \\
&\leq \frac{n^*-1}{2(2\mathrm{e}-1)} \frac{\left( \frac{R^*}{2} \right)^{n^*-1}}{(n^*-1)!} \\
&\leq \frac{n^*-1}{2\mathrm{e}(2\mathrm{e}-1)} \frac{\left( \frac{\mathrm{e}R^*}{2(n^*-1)} \right)^{n^*-1}}{\sqrt{2\pi(n^*-1)}} \\
&\leq \frac{\sqrt{n^*-1}}{2\mathrm{e}(2\mathrm{e}-1)} \frac{\frac{1}{2^{n^*-1}}}{\sqrt{2\pi}} \\
&\leq \frac{1}{2\mathrm{e}(2\mathrm{e}-1)\sqrt{2\pi}} \frac{1}{2^{\frac{n^*-1}{2}}} \\
&\leq \frac{\epsilon}{2}.
\end{aligned}
$$

The second inequality follows from $(n-1)! \geq (n^*-1)!(n^*-1)^{n-(n^*-1)}$. The third and forth inequalities follow from $n^* \geq \mathrm{e}R^* + 1$. In the fifth inequality, we have applied the following lower bound to $(n^*-1)!$, based on Stirling's approximation:

$$
\sqrt{2\pi}(n^*-1)^{n^*-1+0.5} \exp(-n^*+1) < (n^*-1)!.
$$

In the sixth inequality, we have used $n^*-1 \geq \mathrm{e}R^*$. In the seventh inequality, we have used $\sqrt{n} \leq 2^{0.5n}$ for all $n$. In the eighth inequality, we have used $n^* \geq 2\log_2 \frac{2}{2\mathrm{e}(2\mathrm{e}-1)\sqrt{\pi\epsilon}}$.

Substituting from Eqs. (H.36) and (H.41) in Eq. (H.35), we have

$$\left|\hat{h}(x,y;x_0,y_0) - \hat{h}_{n^*}(x,y;x_0,y_0)\right| \leq \epsilon. \tag{H.42}$$

□

# Bibliography


[1] V. K. Madisetti and D. B. Williams. *The Digital Signal Processing Handbook.* CRC in Cooperation with IEEE Press, USA, 1998.

[2] H. J. Weaver. *Applications of Discrete and Continuous Fourier Analysis.* John Wiely and Sons, Inc., New York, 1983.

[3] B. R. A. Nijboer. The diffraction theory of aberrations. *Ph.D. dissertation (University of Groningen, Groningen, The Netherlands, 1942).*

[4] J. Braat, P. Dirksen, and A. J. E. M. Janssen. Assessment of an extended nijboer-zernike approach for the computation of optical point-spread-functions. *J. Opt. Soc. Am. A*, 19:858–870, 2002.

[5] A. J. E. M. Janssen. Extended nijboer-zernike approach for the computation of optical point-spread functions. *J. Opt. Soc. Am. A*, 19:849–857, 2002.

[6] J. Braat, P. Dirksen, A. J. E. M. Janssen, and A. S. van de Nes. Extended nijboer-zernike representation of the vector field in the focal region of aberrated high-aperture optical system. *J. Opt. Soc. Am. A*, 20:2281–2292, 2003.

[7] A. J. E. M. Janssen, J. J. M. Braat, and P. Dirksen. On the computation of the nijboer-zernike aberration integrals at arbitrary defocus. *J. of Modern Opt.*, 51:687–703, 2004.

[8] P. E. X. Silveira and R. Narayanswamy. Signal-to-noise analysis of task-based imaging systems with defocus. *Appl. Opt.*, 45:2924–2934, 2006.

[9] Saeed Bagheri, Paulo E. X. Silveira, Ramkumar Narayanswamy, and Daniela Pucci de Farias. Design and optimization of the cubic phase pupil for the extension of the depth of field of task-based imaging systems. *Optical Information Systems IV, Bahram Javidi, Demetri Psaltis, H. John Caulfield, Proc. of SPIE*, 6311, 2006.

[10] R. Narayanswamy, P. E. X. Silveira, H Setty, V. P. Pauca, and J. van der Gracht. Extended depth-of-field iris recognition system for workstation environment. *in Proc. SPIE*, 5779, 2005.

[11] A. Castro and J. Ojeda-Castaeda. Asymmetric phase masks for extended depth of field. *Appl. Opt.*, 43:3474–3479, 2004.

[12] J. Ojeda-Castaneda, J. E. A. Landgrave, and H. M. Escamilla. Annular phase-only mask for high focal depth. *Submitted to COSI, OSA Topical Meeting, 2007*, Opt. Lett.:1647–1649, 2005.

[13] Saeed Bagheri, Paulo E. X. Silveira, Ramkumar Narayanswamy, and Daniela Pucci de Farias. Analytical optimal solution of the extension of the depth of field using cubic phase wavefront coding. Manuscript under preparation.

[14] Saeed Bagheri, Paulo E. X. Silveira, and Daniela Pucci de Farias. A novel approximation for the defocused modulation transfer function of a cubic-phase pupil. *Submitted to COSI, OSA Topical Meeting, 2007.*

[15] R. Narayanswamy, A. E. Baron, V. Chumachenko, and A. Greengard. Applications of wavefront coded imaging. *Computational Imaging II, C. A. Bouman and E. L. Miller, eds., Proc. SPIE*, 5299:163174, 2004.

[16] J. Hall. F-number, numerical aperture, and depth of focus. *Encyclopedia of Optical Engineering (Marcel Dekker, Inc)*, pages 556–559, 2003.

[17] E. R. Dowski and W. Thomas Cathey. Extended depth of field through wave-front coding. *Appl. Opt.*, 34:1859–1866, 1995.

[18] W. Thomas Cathey and Edward R. Dowski. New paradigm for imaging systems. *J. App. Opt.*, 41:6080–6092, 2002.

[19] K. Kubala, Edward R. Dowski, and W. Thomas Cathey. Reducing complexity in computational imaging systems. *Optics Express*, 11:2102–2108, 2003.

[20] R. Narayanswamy, G. E. Johnson, P. E. X. Silveira, and Hans B. Wach. Extending the imaging volume for biometric iris recognition. *J. App. Opt.*, 44:701–712, 2005.

[21] E. R. Dowski and G. E. Johnson. Wavefront coding: A modern method of achieving high performance and/or low cost imaging systems. *Current Developments in Optical Design and Optical Engineering VIII, R. E. Fischer and W. J. Smith, eds., Proc. SPIE*, 3779:137–145, 1999.

[22] A. W. Lohmann, R. G. Dorsch, D. Mendliovic, Z. Zalevsky, and C. Ferreira. Space-bandwidth product of optical signal and systems. *J. Opt. Soc. Am. A*, 13:470–473, 1996.

[23] R. Piestun and A. B. Miller. Electromegnetic degrees of freedom of an optical system. *J. Opt. Soc. Am. A*, 17:892–902, 2000.

[24] G. E. Johnson, P. E. X. Silveira, and Edward R. Dowski. Analysis tools for computational imaging systems. *Visual Information Processing XIV, Z. Rahman, R. A. Schowengerdt and S. E. Reichenbach, eds., Proc. SPIE*, 5817:34–44, 2005.

[25] Saeed Bagheri, Daniela Pucci de Farias, George Barbastathis, and Mark A. Neifeld. Reduced-complexity representation of the coherent point-spread function in the presence of aberrations and arbitrarily large defocus. *J. Opt. Soc. Am. A*, 23:2476–2493, 2006.

[26] (Invited Paper) Saeed Bagheri, Daniela Pucci de Farias, George Barbastathis, and Mark A. Neifeld. On the computation of the coherent point-spread function using a low-complexity representation. *Optical Information Systems IV, Bahram Javidi, Demetri Psaltis, H. John Caulfield, Proc. of SPIE*, 6311, 2006.

[27] Saeed Bagheri, Paulo E. X. Silveira, and Daniela Pucci de Farias. Limit of signal-to-noise ratio improvement in the imaging systems with extended depth of field. Manuscript under preparation.

[28] Saeed Bagheri, Paulo E. X. Silveira, and Daniela Pucci de Farias. The maximum extension of the depth of field of snr-limited wavefront coded imaging systems. *Submitted to COSI, OSA Topical Meeting, 2007.*

[29] K. H. Brenner, A. W. Lohmann, and J. Ojeda-Castaneda. The ambiguity funcation as a polar display of the otf. *J. Opt. Communications*, 44:323–326, 1983.

[30] E. B. Eliezer, Z. Zalevsky, E. Marom, and N. Konforti. All-optical extended depth of field imaging system. *J. Opt. A: Pure Appl. Opt.*, 5:164–169, 2003.

[31] M. Born and E. Wolf. *Principles of Optics.* Cambridge University Press, UK, 1992.

[32] B. H. W. Hendriks, J. J. H. B. Schleipen, S. Stallinga, and H. van Houten. Optical pickup for blue optical recording at na=0.85. *Opt. Rev.*, 6:211–213, 2001.

[33] H. P. Urbach and D. A. Bernard. Modeling latent-image formation in the photolithography, using the helmholtz eq. *J. Opt. Soc. Am. A*, 6:1343–1356, 1989.

[34] H. A. Buchdahl. *Optical Aberration Coefficients.* Oxford University Press, London, 1958.

[35] J. W. Goodman. *Introductions to Fourier Optics.* McGraw Hill Companies, Boston, 1996.

[36] A. Papoulis. *Signal Analysis.* McGraw-Hill., New York, 1977.

[37] P. M. Woodward. *Probability and Information Theory, with Applications to Radar.* Artech House, Inc., Dedham, Massachusetts, 1980.

[38] A. Papoulis. Ambiguity function in fourier optics. *J. Opt. Soc. Am.*, 64:779–788, 1974.

[39] J. Ojeda-Castaneda and E. E. Sicre. Bilinear optical systems wigner distrbution function and ambiguity representation. *OPTICA ACTA*, 31:255–260, 1984.

[40] H. M. Ozaktas, Z. Zalevsky, and M. A. Kutay. *The Fractional Fourier Transform.* John Wiley and Sons, UK, 2001.

[41] P. S. R. Diniz. *Adaptive Filtering, Algorithms and Practical Implementation.* Kluwer Academic Publishers, Boston, 1997.

[42] A. Yariv. *Optical Electronics.* Saunders College Publishing, Phiadelphia, 1991.

[43] A. Yariv. *Quantum Electronics.* John Wiley and Sons, Unites States, 1989.

[44] A. Mattuck. *Introduction to Analysis.* Prentice Hall, Inc., New Jersey, 1999.

[45] C. L. Tranter. *Bessel Functions with Some Physical Applications.* Hart Pub. Co., New York, 1969.

[46] A. Gray and G. B. Mathews. *A Treatise on Bessel Functions and Their Applications to Physics.* Dover Pub. Inc., New York, 1966.

[47] G. N. Watson. *A Treatise on the Theory of Bessel Functions.* Cambridge University Press, London, 1944.

[48] L. Landau. Monotonicity and bounds on bessel functions. *Mathematical Physics and Quantum Field Theory, Proc. of, H. Warchall, eds.*, 04:147–154, 2000.